Cartwright, Edward

Working Paper

On the emergence of social norms

Department of Economics Discussion Paper, No. 07,04

Provided in Cooperation with:
University of Kent, School of Economics

Suggested Citation: Cartwright, Edward (2007) : On the emergence of social norms, Department of Economics Discussion Paper, No. 07,04, University of Kent, Department of Economics, Canterbury

This Version is available at:
https://hdl.handle.net/10419/68135

On the Emergence of Social Norms

Edward Cartwright
Department of Economics,
Keynes College,
University of Kent,
Canterbury, Kent.
CT2 7NP. UK.
E.J.Cartwright@kent.ac.uk

March 2007

.

Keywords: Social norms, conformity, best reply
JEL classification: C7, D11

Abstract

We consider a model of conformity that permits a non-conformist equilibrium and multiple conformist equilibria. Agents are assumed to behave according to a best reply learning dynamic. We detail the conditions under which a social norm and conformity emerge. The emergence of conformity depends on the distribution of intrinsic preferences, the relative costs and benefits of conformity and the topology of agent interaction.

1 Introduction

Social norms are pervasive in human society. For example, consumption norms dictate what clothes, cars or music are 'acceptable' or 'fashionable' while work norms dictate 'proper effort', 'normal working hours' or a 'reasonable wage' (Lewis 1967, Akerlof 1980, Jones 1984, Elster 1989, Bernheim 1994). Norms can be sustained because of 'costs to non-conformity' in the form of indirect costs, such as guilt, or more direct costs, such as being forced out of employment (Kreps 1997). But: Why do some actions become norms but not others? Why do norms emerge in some choice settings and not others? and, Can policy makers influence behavior though manipulating norms?[1] In this paper we study a simple multiple equilibrium model of conformity with the aim of gaining some insight on these issues.

In the model, an agent's payoff is a sum of intrinsic utility, determined solely by his own action, and social utility, determined by how his action 'fits with that of others'. Social utility is determined relative to some norm of behavior and depends on how closely an agent's actions conform to the norm and on the proportion of the population that are conforming to the norm. If nobody conforms then social utility is zero irrespective of behavior and agents maximize utility by maximizing intrinsic utility. If everybody conforms to the norm then social utility is such that all agents maximize utility by conforming, even if they sacrifice intrinsic utility in doing so. This implies multiple equilibria, including a 'non-conformist' equilibrium where all agents maximize intrinsic utility (or 'do what they want') and 'conformist equilibria' where all agents conform to some norm and receive maximal social utility (but sacrifice intrinsic utility).

What equilibrium should we expect to emerge? To address this equilib-

[1] The Fresno State Social Norms Project is one example of attempting to influence behavior through manipulating norms. The aim of the project is to reduce alcohol abuse through changing student perceptions of 'normal behavior'. Other possibilities include influencing attitudes to saving for retirement, recycling, playing truant at school etc.

rium selection question we use a standard evolutionary or learning model approach (Fudenberg and Levine 1998). Agents are modelled as interacting repeatedly over time and choosing an action using a best reply rule: an agent chooses the action in this period that would have maximized his payoff in the last period. Occasional shocks perturb the system and thus allow the dynamic to potentially evolve between different conformist equilibria or from a conformist to non-conformist equilibrium and so on. We shall detail the conditions under which conformity can emerge and the actions that may become norms.

Why can an action emerge as a norm? For some agents, conforming to the norm may be a relatively 'easy option' because any sacrifice in intrinsic utility is small and so a little social utility is enough to compensate. As the proportion of conforming agents increases then the social utility from conformity increases and other agents become willing to sacrifice increased amounts of intrinsic utility to conform. Conformity can therefore spread through a ripple effect to those who must sacrifice the most intrinsic utility to conform. This logic can be reversed to argue that non-conformity can emerge because 'not-conforming' is a relatively 'easy option' for those who could gain most intrinsic utility from 'not-conforming', and so on. The balance is tipped towards the emergence of conformity if conforming is the 'easy option' for a larger proportion of agents than is not-conforming. Norms most likely to emerge are those that require relatively little sacrifice in intrinsic utility for a significant proportion of the population.

For the most part we focus on a global interaction setting where agents could be seen to interact on a population wide level. Thus, social utility is determined by the total population. In Section 5 we consider an alternative, local interaction setting, where norms and social utility are determined relative to a particular location. The analysis of the global interaction setting naturally extends to the local interaction setting. The local interaction setting does, however, generate 'richer dynamics' of multiple, evolving norms as

we shall illustrate through examples. These examples will demonstrate how the dynamics of conformity depend on the 'topology of agent interaction'.

We shall not attempt in this paper to explain why individuals have desires for social utility or 'to fit in'.[2] It will be taken as given that a conformist equilibrium exists in the sense that *if* everybody conforms to a norm then everybody would want to conform to the norm. As such, we do not attempt to provide a complete story of why a conformist equilibrium would emerge. By detailing, however, when a conformist or non-conformist equilibrium can emerge and by characterizing the actions that can become norms we can provide important insights on the emergence of norms.

Closely related results are due to Akerlof (1980) and Azar (2004). Azar (2004) questioned why a tipping norm can persist. He showed that for a norm to persist there must be sufficient agents who 'like tipping'. This result is consistent with our analysis where an action becomes a norm if there are sufficiently many agents who receive high intrinsic utility from conforming. Akerlof (1980), using the example of a norm to set artificially high wages, questioned whether social custom would be gradually eroded if it is costly for individuals to persist in the custom. He showed that custom can survive. Our analysis suggests that one could go further by saying that conformity or social custom can not only survive but emerge, even if it is not in agents interests for it to do so. Indeed, whether or not conformity or non-conformity emerges in our model is unrelated to the relative Pareto ranking of conformist and non-conformist equilibria. One strand of the existing literature on conformity attempts to explain social norms as 'optimal' in the sense that a conformist equilibrium Pareto dominates a non-conformist equilibrium (Elster 1989). Our results reflect the common thread in the literature on best reply dynamics that the risk dominance of equilibria and not Pareto ranking

[2]See, amongst others, Jones (1984), Bernheim (1994), Kreps (1997) or just about any social psychology textbook for more on this issue. Also, Wooders, Cartwright and Selten (2006) show that in any game with many agents there exists a 'conformist equilibrium' (where similar agents perform similar actions) irrespective of any desires for social utility.

is key (Young 1993). Related results are also due to Ochs and Park (2004) who model a dynamic adoption process where agents subscribe to a network if network size is sufficiently large. Differences in agent preferences create a ripple effect, similar to the one of this paper, where the more subscribe to the network the more will be induced to subscribe to the network.

The motivation behind this paper was to study 'emotional' or 'social' conformity whereby individuals conform to 'fit in' or 'avoid guilt' etc. In reality social utility, as we define it, could be interpreted much more generally. For example 'social utility' could reflect tangible benefits from agent coordination or institutionalized punishment for not obeying rules. Our framework is, however, somewhat distinct from that modelling conformity in coordination games (e.g. Young 1993) because non-conformity (or non-coordination) is an equilibrium and indeed payoffs at the non-conformist equilibrium may exceed those of the conformist equilibrium. There is also no appeal to incomplete information in our model. This distinguishes our results from the literature on 'informational' conformity (e.g. Bikhandani, Hirshleifer and Welch 1992, Juang 2001) where agents imitate successful or popular actions in the hope of obtaining a higher intrinsic utility.

We proceed as follows: Section 2 introduces the model, Section 3 presents the main results, Section 4 provides examples of payoff functions, Section 5 discusses local interaction and Section 6 concludes. All proofs are contained in an Appendix

2 Model and notation

The model used is inspired by that of Bernheim (1994).[3] There exists a continuum of agents. Each agent has a *type* from a set $T \equiv [-1, 1]$. A *population* is described by a cumulative distribution function F, with cor-

[3]There are some notable differences. In particular, Bernheim (1994) provides a more subtle model of conformity in which actions serve as a signal of type.

responding continuous probability density function f, over the set of types T where $f(-1), f(1) > 0$.[4] Agents simultaneously choose an action from set $X = [-1, 1]$. For simplicity, agents of the same type will be assumed to choose the same action.[5] This allows us to describe actions by an *action profile* a that maps T into X with $a(t)$ indicating the action chosen by agents of type t. Let A denote the set of action profiles.

The payoff of an agent is the sum of two components - *intrinsic utility* and *social utility*. We define each in turn.

2.1 Intrinsic utility

Intrinsic utility depends on the difference between type and action; formally, there exists function $I : [0, 2] \to \mathbb{R}$ such that an agent of type t receives intrinsic utility $I(|t - x|)$ from choosing action x.[6] We make the following assumption,

Assumption 1: $I(z)$ is continuous, achieves a maximum at $z = 0$ and is (weakly) concave.

Thus, an agent of type t maximizes his intrinsic utility by choosing action $x = t$. The further his chosen action from t then the lower his intrinsic utility. We shall say that *intrinsic utility is linear* if $I(z) = \beta_1 - \beta_2 z$ for some $\beta_2 > 0$.

2.2 Social utility

Social utility will be determined relative to some norm. This means making assumptions of what 'the norm' is. We will assume that an action constitutes the norm if it is being chosen by a larger proportion of the population than any other action. Let $\rho(x, a)$ be the proportion of the population choosing

[4] More formally, we can allow that $f(-1), f(1) = 0$ but require that $f(-1+\alpha), f(1-\alpha) > 0$ for some $\alpha > 0$ arbitrarily small.

[5] See footnote 12 in Section 2.5 for further elaboration on this assumption.

[6] To simplify notation we shall write $I(t - x)$ instead of $I(|t - x|)$.

action x given action profile a.[7] We shall denote by $\mu(a)$ the 'norm' given action profile a. If there exists action $x^* \in X$ such that $\rho(x^*, a) > \rho(x, a)$ for all $x \neq x^*$ then we call x^* the *norm* and set $\mu(a) = x^*$. Otherwise we say that there is *no norm* and set $\mu(a) = \phi$. If there does exist a norm x^* then an agent who chooses x^* is said to *conform* while an agent who does not choose x^* is said to *not conform*.

Given an action profile a where $\mu(a) \neq \phi$ define

$$\Delta(x) := \begin{cases} \frac{\mu(a)-x}{\mu(a)+1} \text{ if } x \in [-1, \mu(a)) \\ \frac{x-\mu(a)}{1-\mu(a)} \text{ if } x \in (\mu, 1(a)] \\ 0 \text{ if } x = \mu(a) \end{cases} . \tag{1}$$

Value $\Delta(x)$ denotes 'relative distance' between action x and the norm $\mu(a)$. The value of Δ can range from 0 to 1 with 0 indicating conformity and 1 that either action 1 or -1, the actions most removed from the norm, are chosen.[8] Let $\rho(a) := \rho(\mu(a), a)$ denote the proportion of the population conforming to the norm. If $\mu(a) = \phi$ then set $\Delta(x) = 0$ for all x and $\rho(a) = 0$.

Social utility will depend on relative distance $\Delta(x)$ and proportion $\rho(a)$; formally, there exists a social utility function $E : [0, 1] \times [0, 1] \to \mathbb{R}$ such that an agent choosing action x given action profile a receives social utility $E(\Delta(x), \rho(a))$. We shall normalize $E(\Delta, 0) = 0$ for all Δ. Thus, social utility is zero irrespective of action if there is no norm. As discussed in the introduction (see also Section 4) 'social utility' is really just a measure of the general 'costs or benefits' to conforming. For example, social utility could reflect network effects or be positive or negative etc.[9] Note also that

[7] In a continuum population $\rho(x, a)$ may equal 0 for all x.

[8] The choice of relative distance (i.e. for $x > \mu$ relative to $1-\mu$ and for $x < \mu$ relative to $\mu-(-1)$) is for convenience and we could obtain equivalent results using absolute distance $x - \mu$.

[9] Kreps (1997) states four reasons for conformity: 1. conformity is costless, 2. conformity permits coordination, 3. conformity is costly but leads to future benefits (e.g. avoidance of guilt), and 4. conformity is desirable in itself. Reason 1 is a reflection of a 'flat' intrinsic utility function. Reasons 2, 3 and 4 could be incorporated within our

social utility does not depend on type or on the norm. This is a simplifying assumption that could be relaxed.

Some justification should be made for our choice of norm. The primary assumption made is that there exists at most one norm. *If* there is only one norm, then modelling the norm as the 'most observed' action seems appropriate but our method of proof would allow us to define the norm differently, for example, as some exogenous focal point. More contentious is the assumption that there be a unique norm given that one may imagine different norms in different sections of the population. On a practical level it is not clear how one would model multiple norms within a social utility function as defined above. Instead, we shall model, in Section 5, multiple norms as arising from local interaction. That is, we shall permit different norms in different sections of the population but use a framework of local interaction to determine what are 'different sections of the population'.

It seems natural that social utility should be decreasing in the 'individual extent of non-conformity' as given by Δ. We allow for the possibility of a discontinuity in Δ at 0 to reflect a possible discrete jump in social utility between conformity and non conformity.

Assumption 2: $E(\Delta, \rho)$ is non-increasing in Δ and continuous in Δ with a possible exception at $\Delta = 0$.

Assumptions 1 and 2 will be made throughout without further acknowledgment. We also introduce two Properties on payoff functions that while harder to interpret shall essentially prove to be necessary and sufficient conditions in the analysis. We discuss both Properties further in Section 4.

Property 1: For any action profile $a \in A$

$$E(0, \rho(a)) - E(\rho(a), \rho(a)) \geq \rho(a) \left[E(0, 1) - E(1, 1)\right]. \tag{2}$$

definition of social utility.

The value $E(0,1)-E(1,1)$ compares conformity with choosing -1 or 1 when all agents conform and thus measures the largest possible differential in social utility. Given an action profile a in which proportion ρ are conforming, Property 1 requires that the drop in social utility from choosing an action x at relative distance of ρ from $\mu(a)$ is equal to ρ times the largest possible drop in social utility. This requires (see Section 4) that social utility drop relatively sharply for non conformity. For example, if claiming zero unemployment benefit is the norm (Akerlof 1980, Lindbeck, Nyberg and Weibull 1999) then Property 1 would require that claiming some unemployment benefit leads to a relatively large drop in social utility.[10] Examples satisfying this property are provided in Section 4. We say that Property 1 *holds with equality* if the inequality in (2) can be replaced by an equality. We say that Property 1 *does not hold* if the inequality in (2) can be replaced by a strictly less than inequality.[11]

2.3 Payoff functions

Payoffs are given by function $u : X \times T \times A \to \mathbb{R}$ where

$$u(x,t,a) = I(t-x) + E(\Delta(x), \rho(a)). \tag{3}$$

is the payoff of an agent of type t from choosing action x given action profile a. We introduce a second Property relating intrinsic utility to social utility.

Property 2: For any action profile a and actions $t, x \in X$ where either $\mu(a) < x < t$ or $t < x < \mu(a)$,

$$[E(0,\delta_t) - E(\delta_x,\delta_t)] - [E(0,\delta_x) - E(\delta_x,\delta_x)] \geq [I(t-x) - I(t-\mu)] - [I(0) - I(x-\mu)] \tag{4}$$

[10] To fit this example with our action space we can equate $x = -1$ with zero unemployment and $x = 1$ with 50 years unemployment. The norm is $\mu = -1$.

[11] This is a strong notion of requiring Property 1 to not hold in that we require the less than inequality to apply for any action profile $a \in A$.

where $\delta_t := \Delta(t)$ and $\delta_x := \Delta(x)$.

If intrinsic utility is linear then the right hand side of (4) is zero and so, given that $\delta_t > \delta_x$, Property 2 merely requires, as intuition would suggest, that the differential in social utility between conforming and non-conforming should be non-decreasing in the proportion of the population conforming (Akerlof 1980, Lindbeck, Nyberg and Weibull 1999). For example, the less people are claiming unemployment benefit the greater the stigma to claiming unemployment benefit. If intrinsic utility is strictly concave then the right hand side of (4) is strictly positive and so we require more. The more concave is intrinsic utility then the wider need grow the gap in social utility between conforming and non-conforming as the proportion conforming increases. We say that *Property 2 does not hold* if the inequality of equation (4) can be replaced by a strictly less than inequality.

2.4 Nash Equilibrium

A *Nash equilibrium* is an action profile a such that $u(a(t), t, a) \geq u(x, t, a)$ for all $x \in X$ and $t \in T$. The assumptions made are sufficient for the existence of a *non-conformist equilibrium* $\overline{a}$ where $\overline{a}(t) = t$ for all $t \in T$ and $\mu(\overline{a}) = \phi$. That is, social utility is zero and an agent of type t chooses $x = t$, maximizing intrinsic utility. We say that there exists a *conformist equilibrium centered on* x^*, denoted a^{x^*}, where $a^{x^*}(t) = x^*$ for all $t \in T$ if

$$I(t - x^*) + E(0, 1) \geq I(t - x) + E(\Delta(x), 1) \tag{5}$$

for all $t, x \in X$. A *conformist equilibrium* a^{x^*} *is said to be not strict* if there exists at least one t and x combination for which equation (5) holds with equality. We say that *conformity is an equilibrium* if a^x is a conformist equilibrium for any $x \in X$. Finally, we say that *conformity is not a strict equilibrium* if a^x is not a strict conformist equilibrium for any $x \in X$.

2.5 Dynamics

Agents interact over an indefinite number of time periods $\tau = 0, 1, 2, ...$ There exists an initial action profile a^0. Let $a^\tau(t)$ denote the action chosen by agents of type t in period τ. We shall consider a variant of a *best reply dynamic* in which each agent primarily chooses the action for the current period that would have maximized his payoff in the previous period (Fudenberg and Levine 1998). Given action profile a an agent of type t has best reply set $\mathcal{B}_t(a) := \{x \in X : u(x, t, a) \geq u(x', t, a) \text{ for all } x' \in X\}$. The set $\mathcal{B}_t(a)$ may contain multiple actions. In this case we assume agents would pick the action closest to conformity. Let[12]

$$\mathcal{BR}_t(a) := \min_{x \in \mathcal{B}_t(a)} \{|x - \mu(a)|\}.$$

Define,

$$\rho^B(x, a) := \int_{t: x = \mathcal{BR}_t(a)} f(y)dy$$

as the proportion of the population for whom x is a best reply given action profile a.

For the most part we assume that the proportion of the population who choose action x in period τ is given by $\rho(x, a^\tau) = \rho^B(x, a^{\tau-1})$. However, in each period τ with probability $\lambda > 0$ there is a *shock*. If there is a shock then some action $x \neq \mu(a)$ is randomly selected and the proportion choosing x is given by[13]

$$\rho(x, a^\tau) = \rho^B(x, a^{\tau-1}) + \varepsilon$$

[12] In the model agents who are indifferent between conforming and not-conforming will conform. This can clearly be criticized for making conformity more likely. In reality, any tie breaking rule would suffice and all of our results would hold with the rule $\mathcal{BR}_t(a) = \min_{x \in \mathcal{B}_t(a)}\{|x - t|\}$. The rule we use has the merits of greatly simplifying the analysis.

[13] Or, if $\rho^B(x, a^{\tau-1}) > 1 - \varepsilon$ then $\rho(x, a^\tau) = 1$.

where $\varepsilon > 0$ is a real number. Further if $x^* = \mu(a^{\tau-1})$ then[14]

$$\rho(x^*, a^\tau) = \rho^B(x^*, a^{\tau-1}) - \varepsilon.$$

We shall use the phrase 'a shock to action x' if x is selected.

The dynamic modelled is one of best reply with shocks where shocks are minor perturbations to the proportion choosing some action and could be equated with experimentation or a response to advertising etc. It is a zero probability event that a 'positive shock' occurs twice for the same action and so the proportion choosing an action purely because of shocks cannot exceed ε. A 'positive shock' is, however, mirrored by a 'negative shock' to the proportion conforming to the norm. Thus, a norm can be subject to repeated negative shocks. This makes it more difficult for sustained conformity to emerge. Our results are not dependent on this specific model of shocks, chosen for convenience. Note that we shall equate a state of the dynamic with an action profile. More formally, the state of the dynamic is given by function ρ detailing the proportion choosing each action.

3 The emergence of conformity

We begin by detailing the conditions under which the proportion conforming to a norm will grow or diminish. If there exists a norm x^* then those agents with types near to x^* have the most incentive to conform. The distribution of types around x^* thus proves fundamental. Let

$$G(\gamma, x^*) = F((1-\gamma)x^* + \gamma) - F((1-\gamma)x^* - \gamma) \tag{6}$$

for all $\gamma \in [0, 1]$. Function G measures how agent types are distributed around x^*. Clearly, $G(0, x^*) = 0$ and $G(1, x^*) = 1$. If F is the uniform distribution then $G(\gamma, x^*) = \gamma$. If, say, $G(0.1, x^*) = 0.9$ then we could say

[14] Or, if $\rho^B(x^*, a^{\tau-1}) < \varepsilon$ then $\rho(x^*, a^\tau) = 0$.

that a high proportion of agents have types near to x^*.

Theorem 1: Consider action profile a where $x^* = \mu(a)$. If Properties 1 and 2 hold and there exists a conformist equilibrium centered on x^* then $\rho^B(x^*, a) \geq G(\rho(a), x^*)$. If intrinsic utility is linear, Property 1 holds with equality, Property 2 holds and conformist equilibrium a^{x^*} is not strict then $\rho^B(x^*, a) = G(\rho(a), x^*)$. If intrinsic utility is linear, there does not exist a strict conformist equilibrium and Property 1 does not hold or Property 1 holds with equality and Property 2 does not hold then $\rho^B(x^*, a) < G(\rho(a), x^*)$.[15]

If conformity to x^* is to increase then we require $\rho^B(x^*, a^\tau)$ to increase. Theorem 1 indicates that this will depend on the distribution of types around the norm and whether Properties 1 and 2 hold. In particular, if proportion ρ^τ are conforming to norm x^* in period τ then Properties 1 and 2 imply (see the Appendix for details) that those agents with types within distance ρ^τ of x^* will conform in period $\tau + 1$. Proportion $G(\rho^\tau, x^*)$ have types within distance ρ^τ of x^*. Thus, $\rho^{\tau+1} = \rho^B(x^*, a^\tau) \geq G(\rho^\tau, x^*)$. If conformity is to be sustained and/or increase then we need $\rho^{\tau+1} \geq \rho^\tau$. So we need that $G(\rho^\tau, x^*) \geq \rho^\tau$. That is we require sufficiently many agents with types near to x^*.

Applying Theorem 1 we first provide conditions such that the non-conformist equilibrium will occur with vanishing frequency in the long run.

Corollary 1: If Properties 1 and 2 hold and conformity is an equilibrium then for sufficiently small ε as λ tends to zero the probability that $\rho(a^\tau) = 0$ for large τ also tends to zero.

Thus, given Properties 1 and 2 conformity will emerge to some extent in that a norm will exist and a positive proportion of the population will sacrifice

[15] Note that the x^* here could be some exogenous focal point. Thus, the result is more genral than assuming the norm is the action played by most agents.

intrinsic utility in order to conform. The proportion who do conform, however, could be small. The following result provides conditions under which a norm will emerge where all agents conform.

Corollary 2: If Properties 1 and 2 hold, conformity is an equilibrium and there exists $x^* \in X$ such that $G(\gamma, x^*) > \gamma$ for all $\gamma \in (0,1)$ then as λ tends to zero the probability that $\rho(a^\tau) \geq 1 - \varepsilon$ for large τ tends to one.

For conformity to increase over time we require that $G(\rho^\tau, x^*) > \rho^\tau$ over consecutive periods. The requirement that $G(\gamma, x^*) > \gamma$ for all $\gamma \in (0,1)$ captures this in ensuring that $\rho^B(x^*, a^\tau) \geq G(\rho^\tau, x^*) > \rho^\tau$. In each period, the increase in the proportion of agents conforming is enough to entice agents with 'more extreme preferences' to conform and so on. Conformity spreads from those with types near to x^* to those with more extreme types. The requirement that $G(\gamma, x^*) > \gamma$ for all $\gamma \in (0,1)$ is relatively mild. For example, it is satisfied if the distribution over types is unimodal.[16] We provide some illustrations after our next result. Corollaries 1 and 2 do not specify what norms will emerge but it is clear that actions for which G is 'large' are the most likely candidates. The following result is immediate from Theorem 1.

Corollary 3: Suppose that intrinsic utility is linear and Property 1 holds with equality. If $\rho(a^\tau) > \varepsilon$ and the conformist equilibrium $a^{\mu(a^\tau)}$ is not strict then $G(\rho(a^\tau), \mu(a^\tau)) \geq \rho(a^\tau)$.

In reality any norm could emerge if the potential gains to esteem are sufficiently large. Corollary 3, however, shows that whether an x^* norm could emerge will depend on G. To illustrate where norms can emerge consider

[16] By unimodal we mean $F(t) \leq 0.5t$ for all $t \in [-1, x^*]$, $F(t) \geq 0.5t$ for all $t \in [x^*, 1]$ and $f(\mu) > f(t)$ for $t \neq \mu$. Here

$$\begin{aligned} G(\gamma, x^*) &= F((1-\gamma)x^* + \gamma) - F((1-\gamma)x^* - \gamma) \\ &> 0.5\left((1-\gamma)x^* + \gamma\right) - 0.5\left((1-\gamma)x^* - \gamma\right) = \gamma. \end{aligned}$$

three examples depicted in Figure 1. For each Example we indicate the range of x^* for which $G(\gamma, x^*) > \gamma$ for all γ.

Peaked at 0: If $f(t) = 1 - |t|$ then $x^* \in (-0.5, 0.5)$.

Peaked at -1: If $f(t) = \frac{1}{2}(1 - t)$ then $x^* < 0$.

Bimodal: If f is symmetric around 0 with $f(t) = \frac{1}{4}$ for $t \in [0, 0.1)$; $f(t) = 4$ for $t \in [0.1, 0.2]$; $f(t) = \frac{15}{64}(1-t)$ for $t \in (0.2, 1]$ then $x^* \in [0.1, 0.2] \cup [-0.2, -0.1]$.

As already stated the requirement that $G(\gamma, x^*) > \gamma$ for all γ does seem mild but it need not hold. If, for example, approximately half of agents have types near to -1 and half near to 1 then the requirement is not met. Conformity need not spread in this case because if, say, all those with types near to -1 conform to an $x^* = -1$ norm social utility need not be such that those agents with types near to 1 wish to conform. Conformity does not spread because of the disparity in tastes.

4 Conformity and social utility

So far we have emphasized the importance of the distribution over types in determining the emergence of conformity. In this section we shall explore the role of payoff functions and the social utility function. If the distribution over types is uniform then $G(\gamma, x) = \gamma$ for all γ and all x and so, given the discussion above, this presents a limiting case where conformity is 'least likely to emerge'. The following result demonstrates the necessity of Properties 1 and 2 in this case.

Corollary 4: Suppose that types are distributed uniformly, conformity is not a strict equilibrium and intrinsic utility is linear. If Property 1 does not hold, or Property 1 holds with equality and Property 2 does not hold then

$\rho(a^\tau) \leq \varepsilon$ for all τ.[17]

If the distribution over types is not uniform (or there exists a strict conformist equilibrium) then Properties 1 and 2 are not necessary; the 'incentives to conform' inherent in the distribution over types could compensate for the 'lack of incentives to conform' implied by a relaxation of Properties 1 and 2. Clearly, however, Properties 1 and 2 are important in determining the emergence of conformity. We consider four examples, illustrated in Figure 2, to discuss the properties. In looking at Property 2 suppose that $I(z) = \beta_1 - \beta_2 z - \beta_3 z^2$ for all z and some real numbers β_1 and $\beta_2, \beta_3 \geq 0$ and $\mu = 0$. Property 2 reduces to $[E(0,t) - E(x,t)] - [E(0,x) - E(x,x)] \geq 2\beta_3 x(t-x)$ for $t > x > 0$ and $t < x < 0$.

Example 1, Conform or not: Consider the simplest form of social utility function

$$E_1(\Delta, \rho) = \begin{cases} 0 \text{ if } \Delta \neq 0 \\ h(\rho) \text{ if } \Delta = 0 \end{cases}$$

for some function h. There is no distinction in 'how close an agent is to conforming'. An agent who conforms receives positive social utility and one who does not receives zero social utility. It is immediate from (2) that Property 1 holds if there is *concavity in social utility with respect to* ρ. That is, if $h(\rho) \geq \rho$ for all ρ. Thus we require social utility to be 'relatively responsive to ρ' when ρ is small. Property 2 requires β_3 to be sufficiently small. For example, if $h(\rho) = \rho$ then $\beta_3 \leq 0.5$ would do.

This form of social utility function is extreme in its distinction between conformity and non-conformity and easily satisfies the two properties but is not unreasonable for certain choice settings. Bernheim (1994) considers a model of conformity where any deviation from a norm is seen as a signal of an agent with 'most extreme types'. Thus, any deviation from the norm leads

[17] By Theorem 1 we know that $\rho^B(x^*, a) < G(\rho(a), x^*)$. Given that f is uniform we have that $G(\rho(a), x^*) = \rho(a)$. The result is now immediate.

to a substantial loss in social utility. Function E_1 may also be appropriate in dealing with network or coordination effects. If there is a positive externality from using the same action as others then this could be modelled by setting $h(\rho) = \rho$.

Example 2, a continuous social utility function: Let,

$$E_2(\Delta, \rho) = -\rho^{\gamma_1}\Delta^{\gamma_2}$$

for all $\Delta \in [0,1]$ and some real numbers $\gamma_1, \gamma_2 > 0$. This social utility function is continuous and so there is no discrete drop in utility for non-conformity. Property 1 does, however, require that social utility fall 'steeply' for deviations from μ. It turns out, that 'steep enough' requires $\gamma_1 + \gamma_2 \leq 1$. Formally,

$$E_2(0,\rho) - E_2(\rho,\rho) = \rho^{\gamma_1}\rho^{\gamma_2} \geq \rho\left[E_2(0,1) - E_2(1,1)\right] = \rho$$

if $\gamma_1 + \gamma_2 \leq 1$ (where $\rho \leq 1$). So, we require *convexity in social utility with respect to* Δ whereby social utility is more responsive to Δ the smaller is Δ. For example, the change from claiming one to two months unemployment benefit results in a relatively large change in stigma compared to a change from eight to nine months benefit.

Example 3, a concave social utility function: For social utility to be concave in Δ we need that differences in social utility grow rapidly for small Δ. The most extreme example of this is,

$$E_3(\Delta, \rho) = \begin{cases} -\Delta^2 \text{ if } \Delta \neq 0 \text{ and } \rho > 0 \\ 1 \text{ if } \Delta = 0 \text{ and } \rho > 0 \end{cases}$$

Recall that $E_3(\Delta, 0) = 0$ by assumption for all Δ. Thus, once a positive proportion of the population conform the full differences in social utility exist. An extreme example but one where concavity in social utility with

respect to Δ is compensated by the responsiveness of social utility to ρ. Property 1 is satisfied because

$$E_3(0,\rho) - E_3(\rho,\rho) = 1 + \rho^2 \geq \rho\left[E_1(0,1) - E_1(1,1)\right] = 2\rho$$

One setting where this example may be appropriate is if there exists an 'institutionalized punishment mechanism' where there exists a norm and 'authority' punishes any deviation from the norm.[18] In this context the punishment need not depend on ρ.

Example 4, summary: Consider a general functional form,

$$E_4(\Delta,\rho) = \begin{cases} \rho(2-\rho)(1-\Delta)^{\gamma_3} \text{ if } \Delta \neq 0 \\ \rho(2-\rho)(1+\gamma_4) \text{ if } \Delta = 0 \end{cases}$$

for some real numbers $\gamma_3, \gamma_4 \geq 0$. To satisfy Property 1 we require that

$$E_3(0,\rho) - E_3(\rho,\rho) = \rho(2-\rho)(1+\gamma_4-(1-\rho)^{\gamma_3}) \geq \rho\left[E_3(0,1) - E_3(1,1)\right] = \rho(1+\gamma_4)$$

implying that

$$(1+\gamma_4)(1-\rho) \geq (2-\rho)(1-\rho)^{\gamma_3}$$

for all $\rho \in (0,1)$. This is satisfied if $\gamma_3 \geq 1$ and $\gamma_4 \geq 1$. That $\gamma_3 \geq 1$ implies that E_4 is (weakly) convex in Δ. That $\gamma_4 \geq 1$ implies that E_4 is discontinuous at $\Delta = 0$. Property 2 requires

$$(1+\gamma_4-(1-x)^{\gamma_3})\left[t(2-t) - x(2-x)\right] \geq 2\beta_3 x(t-x).$$

This is satisfied if $1 + \gamma_4 \geq 2\beta_3$. So, again we require β_3 small.

In summary, a social utility function 'most likely' to satisfy Properties 1 and 2 would, as we see in Example 4, be discontinuous in social utility at $\Delta = 0$

[18] Our definition of a norm (as the most used action) may not be appropriate in this context but the analysis is easily adapted to an exogenous norm.

and have the concavity in ρ and convexity in Δ properties of Examples 1 and 2. If an agent can achieve relatively high social utility through non-conformity, even when the proportion conforming is small, then Property 1 is not satisfied and conformity need not emerge.

Note that the relationship between intrinsic utility and social utility is important in determining whether or not there exists a conformist equilibrium centered on a norm x^*. Clearly, however, if social utility is sufficiently large relative to intrinsic utility and $E(0,1) > E(x,1)$ for all $x \neq x^*$ then there always exists a conformist equilibrium centered on x^*. The 'shape of the social utility function' is also important in determining the behavior of agents who do not conform. In Example 1 an agent of type t who does not conform will clearly choose $x = t$. More generally, an agent of type t, even if he does not conform, may choose an action $x \neq t$ and so is still influenced by social utility.

5 Local interaction

An extension of the model is to allow local interaction (Fudenberg and Tirole 1998). That is, to suppose an agent may interact with only a subset of the population or that his social utility is determined relative to a subset of the population. In many instances this seems appropriate, perhaps reflecting simple geography, such as a norm within a workplace, or that agents only care about the views of certain others, such as friends. Local interaction also has the advantage of permitting 'richer norm dynamics'. In particular, there can be multiple norms and an increased opportunity for norms to change over time. How to model local interaction is, however, not so clear and so we will limit ourselves here to two objectives. First, to illustrate how the previous analysis can be extended to a local interaction setting. Second, to illustrate how 'richer norm dynamics' can emerge with local interaction. We leave a more complete exploration of local interaction to future research.

Suppose that there exists a set of Locations $L = [-1, 1]$. There is a continuous probability density function f over $T \times L$ detailing how agents are distributed across types and locations. Local interaction is characterized by a real number $\eta \in (0, 2]$. An agent interacts with all those agents within distance η of his location. Given an action profile a, action x and location l let $\rho(x, l, a)$ denote the proportion of those agents located between $l - \eta$ and $l + \eta$ who choose action x. Formally, if $a(t, l)$ states the action chosen by those of type t at location l then[19]

$$\rho(x, l, a) := \frac{\int_{l-\eta}^{l+\eta} \int_{t:a(t,y)=x} f(p, y) dp dy}{\int_{l-\eta}^{l+\eta} \int_{t \in T} f(p, y) dp dy}.$$

Given action profile a, if there exists action $x^* \in X$ such that $\rho(x^*, l, a) > \rho(x, l, a)$ for all $x \neq x^*$ then we call x^* the *norm at location* l and set $\mu(l, a) = x^*$. Otherwise we say that there is no norm at location l and set $\mu(l, a) = \phi$. An agent conforms if he is at location l and chooses action $\mu(l, a)$. Note that at different locations there may be different norms.

The social utility function can be defined as previously where Δ and ρ are determined relative to a location. For example, an agent at location l receives social utility $E(\Delta(x), \rho(\mu(l, a), l, a))$ from choosing action x where $\Delta(x)$ measures the relative distance between x and $\mu(l, a)$. If $\eta = 2$ then there is global interaction and we have the situation previously modelled. If $\eta < 2$ then we have local interaction. Extending the dynamic (and notation) in a natural way, if there is a shock in period τ, an action x and location l^* are randomly selected and

$$\rho(x, l, a^\tau) = \rho^B(x, l, a^{\tau-1}) + \varepsilon$$

while

$$\rho(\mu(l, a^{\tau-1}), l, a^\tau) = \rho^B(\mu(l, a^{\tau-1}), l, a^{\tau-1}) - \varepsilon.$$

[19] If $l-\eta$ or $l+\eta$ do not belong to L then truncate the range of the integrals as appropriate.

for all $l \in [l^* - \eta, l^* + \eta]$.

In Section 3 we saw that the value of G as a measure of the distribution of types informed whether or not conformity could invade. We introduce the analogue in a local interaction setting. Fix an action profile a and location l^* where $\mu(l^*, a) = x^*$. For all l let $\rho_l := \rho(x^*, l, a)$ and define $t_l^+ := (1-\rho_l)x + \rho_l$ and $t_l^- := (1-\rho_l)x - \rho_l$. Define,

$$G(a, \mu(l^*, a), l^*) := \frac{\int_{l^*-\eta}^{l^*+\eta} \int_{t_y^-}^{t_y^+} f(p, y) dp dy}{\int_{l^*-\eta}^{l^*+\eta} \int_{-1}^{1} f(p, y) dp dy}.$$

The G function just defined is analogous to that used earlier and serves the same purpose.[20] It gives a measure of the proportion of agents with whom an agent at location l^* interacts who will conform to x^*.

Corollary 5: If Properties 1 and 2 hold and conformity is an equilibrium then $\rho^B(\mu(l, a), l, a^\tau) \geq G(a^\tau, \mu(l, a), l)$ for all l. If intrinsic utility is linear, Property 1 holds with equality, Property 2 holds and conformist equilibrium $a^{\mu(l,a)}$ is not strict then $\rho^B(\mu(l, a), l, a^\tau) = G(a^\tau, \mu(l, a), l)$.

Given G and Corollary 5 we can begin to explore how conformity may emerge. Two sketch examples allow us to illustrate some of the issues.

Example 5: In this example the emergence of conformity is less likely than with global interaction. Conformity will 'spread' across locations but not necessarily across types.

Let $f(t, l) = 0.25$ for all t and l. Thus, agents are distributed uniformly across locations and types. Suppose that $\mu(a^0, l) = \phi$ for all l and a shock occurs in period 1 at location $l^* = 0$ to choose action $x^* = 0$. Thus $\rho(x^*, l, a^1) = \varepsilon$ for all $l \in [-\eta, \eta]$ and $\rho(x^*, l, a^1) = 0$ for all other l. In period 2 agents at

[20] Recall that $G(\gamma, \mu) = F((1-\gamma)\mu + \gamma) - F((1-\gamma)\mu - \gamma)$. In calculating $G(a, \mu(l^*, a), l^*)$ we are essentially summing the $G(\rho_l, \mu)$ over locations near to l^*. If $\eta = 2$ then $G(a, \mu(l^*, a), l^*) = G(\rho(a), a)$.

locations $l \in [-\eta, \eta]$ who have types 'close' to 0 will conform as illustrated in Figure 3a. Suppose that 'close' is $t \in [-\gamma_\varepsilon, \gamma_\varepsilon]$ for some γ_ε. Now $\rho(x^*, l, a^2) > 0$ and $\mu(l, a^2) = x^*$ for all $l \in (-2\eta, 2\eta)$. The closer, however, is l to l^* then the larger is $\rho(x^*, l, a^2)$. Specifically, $\rho(x^*, l, a^2) = \gamma_\varepsilon(1 - |l|/2\eta)$ for all $l \in (-2\eta, 2\eta)$. The larger is $\rho(x^*, l, a^2)$ then the wider the range of types at location l for which conformity is a best reply. So, in Period 3 we get a 'diamond' of type, location combinations within which agents would conform to x^*. At the limit of the diamond only those with types $t = x^*$ choose x^*. In period 3 we have $\rho(x^*, l, a^3) > 0$ for all $l \in (l^* - 3\eta, l^* + 3\eta)$ while $\rho(x^*, l, a^3)$ is still larger the closer is l to x^*, and so on.

If conformity is to emerge then the 'diamond' of Figure 3a must expand rather than contract over time. From Corollary 5 we know that if $G(x^*, l, a^\tau) > \rho(a^\tau, x^*, l)$ then conformity can spread at location l. As we have seen, however, the further is l from x^* then the lower is $G(x^*, l, a^\tau)$. This makes it more difficult for conformity to emerge in the local interaction setting. Specifically, we know that if intrinsic utility is linear, Property 1 holds with equality and conformist equilibrium a^{x^*} is not strict then $\gamma_\varepsilon = \varepsilon$. Calculation yields $G(x^*, 0, a^2) = \frac{3}{4}\varepsilon$ and $G(x^*, \eta, a^2) = \frac{\varepsilon}{2}$. Crucially, $\rho^B(x^*, l, a^2) = G(x^*, l, a^2) < \varepsilon$ for all l and so conformity is disappearing. We know, however, from Theorem 1, that conformity would emerge in this setting if $\eta = 2$. If $\eta < 2$ then 'twice the density' of agents with types near x^* are required so that $\rho(x^*, -\eta, a^2) \geq \varepsilon$ and conformity can emerge.

Finally, note that, *if* the norm emerges, then $\rho(a^\tau, x^*, l) > 0$ for all $l \in (l^* - \tau\eta, l^* + \tau\eta)$ and so, in the absence of further shocks, the norm will necessarily spread until $\rho(a^\tau, x^*, l) > 0$ for all $l \in L$.■

Example 6: In this example local interaction makes it easier for conformity to emerge. Conformity may not 'spread' across locations.

Suppose that distribution f is such that 'most agents at location l have types near to $t = l$'. That is, location is strongly correlated with type. Agents who

interact therefore tend to have the same type. With the aid of Figure 3b we can trace through the differences with the earlier example. In period 2 the diagram is the same. If, however, η is small then $G(x^*, l^*, a^2) \simeq 1$ and $G(x^*, l^* - \eta, a^2) \simeq 0.5$ because almost all of the agents around l^* have types near to x^*. In terms of Figure 3a all agents at locations around l^* have types 'in the shaded area'. This results in a 'stretching out' of the diamond in Period 3, as illustrated in Figure 3b, because any type of agent would conform with such a high proportion of agents conforming. Consequently $G(x^*, \eta, a^3) \simeq 1$ because all of agents around $l^* - \eta$ will have types near to l^* and in the shaded area, and so on.

Conformity can emerge very easily in this setting. Indeed, any action could become a norm. Even if there are relatively few agents with a particular type x^* this need not stop x^* emerging as a norm provided that an agent of type x^* interacts with other agents who have types near to x^*. Conformity need not, however, spread across locations. That half of the agents with whom a person interacts conform is a large inducement to conform. But, the further removed is l from x^* the fewer are the agents of the type that want to conform even with this inducement. Generally, there exists a set of types $Tn(x^*)$ where the best reply is to not conform even if $\rho(x^*) = 0.5$. At locations distant from l^* it may be that all agents have types belonging to $Tn(x^*)$. Thus, there may be a 'stable state' where an x^* norm exists at locations around l^* but not at more distant locations.■

These two examples illustrate some of the issues concerning the emergence of norms in local interaction setting. We have not, as yet, however, addressed the possible consequences of different norms at different locations. In Example 6 the presence of multiple norms appears inevitable while in Example 5 we might expect that it depends on the frequency of shocks and speed with which conformity spreads. Clearly, however, multiple norms are possible. Suppose that there exists location l such that $\mu(a, l') = x^1$ for $l' < l$ and $\mu(a, l'') = x^2$ for $l'' > l$. A first observation is that $\mu(a, l) = \phi$ and

$\rho(a, x^1, l) = \rho(a, x^2, l)$.[21] Thus, an agent 'at the boundary' between norms must interact with as many conforming to x^2 as conforming to x^1. Revisiting Example 6 allows us to illustrate the implications of this.

Example 6': The distribution over types is assumed to be the same as Example 6. In addition, if $l > l'$ then there are proportionally more agents around location l than location l'. Consider the boundary between norms at location l. If $\rho(a, x^1, l) = \rho(a, x^2, l)$ then proportionally more agents must be conforming to norm x^1 than x^2 to balance up the fact that there are potentially more agents to conform to norm x^2 than x^1. This requires a precise balancing act of forces such as in Figure 4. This may suggest that a 'stable' boundary between norms is unlikely but, on reflection, there is no reason to suppose it should not occur. In particular, at lower locations the norm x^2 will begin to 'die out' and at higher locations the norm x^1 will begin to 'die out' so a stable boundary location may occur. If it does occur, however, norms must be sufficiently distinct. Let $Tc(x^2)$ be the set of types that would want to conform to an x^2 norm if $\rho(x^2) \geq 0.5$. If all agents at location l have types belonging to set $Tc(x^2)$ and $l < x^2$ then, given the bias in f, we should expect $\rho(a, x^2, l) > 0.5 > \rho(a, x^1, l)$ and the norm at l must be x^2. Thus, if $x^1 \in Tc(x^2)$ we could not expect to observe an x^2 and an x^1 norm coexisting. For sufficiently low l, however, some agents will not have types belonging to $Tc(x^2)$ and the x^2 norm will begin to die out. At this point a different norm could exist. Given the correlation between type and location this implies that x^1 and x^2 must be sufficiently distinct.■

This section has only briefly touched on the issues that arise with local interaction but has hopefully served its purpose. One conclusion that we can draw is how the emergence of conformity will depend on the nature of agent interaction. For example, if an agent's social utility is determined by his interaction with agents that have similar preferences then this encourages the

[21] This follows from the definitions of a norm.

emergence of conformity. This also allows multiple norms to emerge across agents with different preferences.

6 Concluding remarks

We have analyzed a model of conformity in which there exists both conformist and non-conformist equilibria and have provided conditions under which conformity does or does not emerge. Three things prove important in the emergence of conformity. First, the distribution of types in the population. The more agents have types 'close to the norm' then the more likely is conformity to emerge. Second, the 'shape of the social utility function'. The larger are the losses in social utility for deviation from the norm then the more likely is conformity to emerge. Finally, the topology of agent interaction. Whether agents interact with subsets of the population and whether they interact with agents of similar types can impact on the emergence of conformity.

We highlight four areas for future research. The issue of local interaction is clearly one area that could be explored more deeply with differing assumptions on the 'network of interaction'. A second issue is the rate at which conformity may emerge. The current paper details the 'long run' outcomes of the dynamic but the literature on equilibrium selection has highlighted the importance of looking also at rates of convergence (Fudenberg and Levine 1998). Intuitively one could expect a short waiting time for conformity to emerge given that its emergence requires only one shock. It may take longer, however, for population wide conformity to emerge. In particular, if a norm and stable state emerge where half of the population conform to some norm it appears unlikely that the norm would change in 'normal time' to one where all agents would want to conform, even if such a norm exists. A related issue is the stability of equilibrium. We have taken the approach in this paper that the initial state could be the non-conformist equilibrium and have provided

conditions under which conformity may emerge. A distinct approach would be to set the initial state as a conformist equilibrium. One would expect that conformity can 'survive' under more general conditions than it can emerge. One final issue is how things change with incomplete information about the norm. For example, if the norm is to tip 15% in a restaurant how should it be interpreted if an agent tips 14%. In the Examples of Section 4 we assumed that social utility dropped quickly for deviations from the norm. This is consistent with Bernheim (1994). Azar (2004), however, makes the case that social utility should not drop quickly around the norm given the 'fuzziness' at the margin as to whether agents are conforming or not. In our framework this argument is not so appropriate given that we also judge the strength of conformity by the proportion choosing the norm (and not something near the norm). It would be interesting, however, to see how the dynamics of conformity change if agents who choose actions near to the norm are seen as potentially conforming, both in terms of the social utility they receive and the perceived level of conformity within in the population.

7 Appendix

Theorem 1 and the corollaries are proved with the help of 6 Lemmas.

Lemma 1: For any action profile a and type t, if $x = \mathcal{BR}_t(a)$ and $t < \mu$ then $t \leq x \leq \mu$ or if $t > \mu$ then $t \geq x \geq \mu$.

Proof: Set $t < \mu$. If $x < t$ then $I(0) \geq I(t-x)$ and $E(\Delta(t), \rho(a)) \geq E(\Delta(x), \rho(a))$. But then $u(t,t,a) \geq u(x,t,a)$ and so $t \in \mathcal{B}_t(a)$ contradicting that $x = \mathcal{BR}_t(a)$. If $x > \mu$ then $I(\mu - t) \geq I(x-t)$ and $E(0, \rho(a)) \geq E(\Delta(x), \rho(a))$. But then $u(\mu, t, a) \geq u(x,t,a)$ and so $\mu \in \mathcal{B}_t(a)$ again contradicting that $x = \mathcal{BR}_t(a)$. A symmetric argument treats $t > \mu$.■

Lemma 2: For any action profile a where $\mu := \mu(a)$ and any types $t_1, t_2 \in T$,

if $\mu = \mathcal{BR}_{t_1}(a)$ and $\mu = \mathcal{BR}_{t_2}(a)$ then $\mu = \mathcal{BR}_t(a)$ for all $t \in (t_1, t_2)$. Also $\mu = \mathcal{BR}_\mu(a)$.

Proof: First, $I(0) \geq I(x)$ for all x and $E(0, \rho) \geq E(\Delta(x), \rho)$ for all ρ and x. Thus, $\mu = \mathcal{BR}_\mu(a)$. If $\mu = \mathcal{BR}_t(a)$ then

$$I(t - x) - I(t - \mu) \leq E(0, \rho(a)) - E(\Delta(x), \rho(a))$$

for all x. Suppose that $t_1 < t < \mu$. By concavity of I

$$\frac{I(t_1 - x) - I(t_1 - \mu)}{\mu - x} \geq \frac{I(t - x) - I(t - \mu)}{\mu - x} \geq \frac{I(\mu - x) - I(0)}{\mu - x}$$

for all $x \geq t_1$ and so $\mu = \mathcal{BR}_t(a)$ whenever $\mu = \mathcal{BR}_{t_1}(a)$. A symmetric argument treats $t_1 > t > \mu$.■

Some notation: Given an action profile a let $t^+ := (1 - \rho(a))\mu(a) + \rho(a)$ and $t^- := (1 - \rho(a))\mu(a) - \rho(a)$.

Lemma 3: Consider action profile a and norm $\mu = \mu(a)$. If Property 1 holds and there exists a conformist equilibrium centered on μ then $u(\mu, t^+, a) \geq u(t^+, t^+, a)$ and $u(\mu, t^-, a) \geq u(t^-, t^-, a)$. If intrinsic utility is linear, Property 1 holds with equality and conformist equilibrium a^μ is not strict then $u(\mu, t, a) < u(t, t, a)$ for any $t < t^-$ or $t > t^+$. If intrinsic utility is linear, there does not exist a strict conformist equilibrium a^μ and Property 1 does not hold then $u(\mu, t^+, a) < u(t^+, t^+, a)$ and $u(\mu, t^-, a) < u(t^-, t^-, a)$.

Proof: Consider t^+. If there exists a conformist equilibrium centered on μ then

$$I(1 - \mu) + E(0, 1) \geq I(0) + E(1, 1). \tag{7}$$

By property 1

$$E(0, \rho(a)) - E(\rho(a), \rho(a)) \geq \rho(a) \left[E(0, 1) - E(1, 1)\right] \tag{8}$$

Note that $\Delta(t^+) = \rho(a)$ and so

$$u(\mu, t^+, a) = I(t^+ - \mu) + E(0, \rho(a)) \geq I(0) + E(\rho(a), \rho(a)) = u(t^+, t^+, a)$$

if

$$\rho(a) \geq \frac{I(0) - I(t^+ - \mu)}{I(0) - I(1 - \mu)}. \tag{9}$$

By Assumption 1 and concavity,

$$I(t^+ - \mu) \geq \left[\frac{t^+ - \mu}{1 - \mu}\right] I(1 - \mu) + \left[1 - \frac{t^+ - \mu}{1 - \mu}\right] I(0)$$

implying that

$$\frac{t^+ - \mu}{1 - \mu} \geq \frac{I(0) - I(t^+ - \mu)}{I(0) - I(1 - \mu)}$$

But, by construction,

$$\frac{t^+ - \mu}{1 - \mu} = \rho(a).$$

This demonstrates the first part of the proof. If intrinsic utility is linear, Property 1 holds with equality and conformist equilibrium a^μ is strict then the inequality of equations (9), (8) and (7) can be replaced with equalities. Thus $u(\mu, t^+, a) = u(t^+, t^+, a)$. For an agent of type $t > t^+$ this implies that $u(\mu, t, a) < u(t, t, a)$. If intrinsic utility is linear, there does not exist a strict conformist equilibrium and Property 1 does not hold then equation (9) holds with equality, (7) with a less than inequality and (8) with a strictly less than inequality implying that $u(\mu, t^+, a) < u(t^+, t^+, a)$ as desired. A symmetric argument treats t^-.■

Lemma 4: Consider action profile a where $\mu := \mu(a)$. If Properties 1 and 2 hold and there exists a conformist equilibrium a^μ then

$$\left[t^-, t^+\right] \subseteq \{t \in T : \mu = \mathcal{BR}_t(a)\}$$

If intrinsic utility is linear, there does not exist a strict conformist equilibrium

a^μ and Property 1 holds with equality but Property 2 does not hold then $\mu \neq \mathcal{BR}_{t^+}(a), \mathcal{BR}_{t^-}(a)$.

Proof: From Lemma 2 we only need check that $\mu = \mathcal{BR}_{t^+}(a), \mathcal{BR}_{t^-}(a)$. Consider t^+. We need to show that

$$I(t^+ - \mu) + E(0, \rho(a)) \geq I(t^+ - x) + E(\Delta(x), \rho(a)) \tag{10}$$

for all x. From Lemmas 1 and 3 we can reduce this to $x \in (\mu, t^+)$. Fix an $x \in (\mu, t^+)$. Let δ be such that $(1-\delta)\mu + \delta = x$. Note that $\delta = (x-\mu)/(1-\mu) = \Delta(x) \leq \rho(a)$ and $\rho(a) = \Delta(t^+)$. So, applying Lemma 3 we know that

$$I(x - \mu) + E(0, \delta) \geq I(0) + E(\Delta(x), \delta). \tag{11}$$

By Property 2

$$I(t^+-\mu)-I(t^+-x)+E(0,\rho(a))-E(\Delta(x),\rho(a)) \geq I(x-\mu)-I(0)+E(0,\delta)-E(\Delta(x),\delta) \tag{12}$$

Combining (11) and (12) gives (10) and the desired result for the first part of the Lemma. Now, suppose that intrinsic utility is linear, Property 1 holds with equality and there does not exist a strict conformist equilibrium a^μ. By Lemma 3, the inequality of (11) becomes a less than inequality. If Property 2 does not hold then the inequality of (12) becomes a strictly less than inequality. Thus, $u(\mu, t^+, a) < u(x, t^+, a)$ for all $x \in (\mu, t^+)$ and so $\mu \neq \mathcal{BR}_{t^+}(a)$ as desired. A symmetric argument treats t^-.■

Proof of Theorem 1: The first part of the statement is immediate from Lemma 4 and the definitions of G, t^+ and t^-. The second part of the statement is immediate from Lemmas 3 and 4 implying that $\{t \in T : \mu = \mathcal{BR}_t(a)\} = [t^-, t^+]$. The final part of the statement is immediate from from Lemmas 2, 3 and 4 implying that $\{t \in T : \mu = \mathcal{BR}_t(a)\} \subset (t^-, t^+)$.■

Lemma 5: For any action profile a and action $x \neq \mu(a)$, if $x = \mathcal{BR}_t(a)$ for

some t then $x \notin \mathcal{BR}_{t'}(a)$ for all $t' \neq t$.[22]

Proof: Suppose otherwise. Thus, $x = \mathcal{BR}_{t'}(a)$ and $x = \mathcal{BR}_{t''}(a)$ for some $t', t'' \in T$ and $x \neq \mu$. Suppose that $t'' \leq \mu$. By Lemma 1, $t' < t'' \leq x < \mu$. Repeating the argument of Lemma 2, $x = \mathcal{BR}_t(a)$ for all $t \in [t', t'']$. Let ν be a small positive number. Given that $x \in \mathcal{BR}_{t''}(a)$

$$I(x - t'') + E(\Delta(x), \rho(a)) > I(x + \nu - t'') + E(\Delta(x - \nu), \rho(a)).$$

By Assumptions 1 and 2 (continuity) this would imply that for ν sufficiently small,

$$I(x - t'') - I(x + \nu - t'') > E(\Delta(x), \rho(a)) - E(\Delta(x + \nu), \rho(a)).$$

But this implies for an agent of type $t = t'' - \nu$ that

$$I(x - \nu - t) + E(\Delta(x + \nu), \rho(a)) > I(x - t) + E(\Delta(x), \rho(a))$$

contradicting that $x = \mathcal{BR}_t(a)$. A symmetric argument treats $t'' > \mu$.■

Lemma 6: Consider action profile a^0 where $\mu(a^0) = x^*$. If Properties 1 and 2 hold, there exists conformist equilibrium a^{x^*} and $\lambda = 0$ then either $\rho(a^0) \leq \rho(a^1) \leq \rho(a^2) \leq ...$ or $\rho(a^0) \geq \rho(a^1) \geq \rho(a^2) \geq ....$ If $G(\gamma, x^*) > \gamma$ for all $\gamma \in (0, \overline{\gamma})$ for some $\overline{\gamma}$ then $\lim_{\tau \to \infty} \rho(a^\tau) > \overline{\gamma}$.

[22] If there is not continuity in the esteem or intrinsic utility function then Lemma 5 may not apply. To illustrate let,

$$E_5(\Delta, \rho) = \begin{cases} 5\rho \text{ if } \Delta = 0 \\ 4\rho \text{ if } \Delta \leq \rho \\ 0 \text{ if } \Delta > \rho \end{cases}$$

and set $I(t - x) = -5(t - x)^2$ and $F(t) = 0.5(1 + t)$. Consider initial state a^0 where $\rho(0, a) = 0.2$ and $\mu(a) = 0$. Now, $E_5(0, \rho) = 1$ and $E_5(0.2, \rho) = 0.8$. Calculation yields, $\mathcal{BR}_t(a) = 0$ for all $t \in [-0.2, 0.2]$ and $\mathcal{BR}_t(a) = 0.2$ for all $t \in (0.2, 0.6]$. Thus $\rho(0, a^1) = \rho(0.2, a^1) = \rho(-0.2, a^1) = 0.2$.

Proof: By Lemma 5, either $\mu(a^\tau) = x^*$ or $\mu(a^\tau) = \phi$. By Lemma 2, to each τ we can associate types t_τ^L and t_τ^H such then $a^\tau(t) = x^*$ for all $t \in (t_\tau^L, t_\tau^H)$ and $a^\tau(t) \neq x^*$ for all other τ. Thus, $\rho(a^\tau) = F(t_\tau^H) - F(t_\tau^L)$. Clearly, $t_{\tau+1}^L \leq t_\tau^L$ if and only if $t_{\tau+1}^H \geq t_\tau^H$. So, if $t_{\tau+1}^L \leq t_\tau^L$ then $\rho(a^{\tau+1}) \geq \rho(a^\tau)$ implying that $t_{\tau+2}^L \leq t_\tau^L$. Iterating the argument gives that $t_{\tau'}^L \leq t_\tau^L$ for all $\tau' > \tau$. Thus, t'_τ is non-increasing in τ. If $t_{\tau+1}^L \geq t_\tau^L$ then $t_{\tau+1}^H \leq t_\tau^H$. So, if $t_{\tau+1}^L \geq t_\tau^L$ then $\rho(a^{\tau+1}) \leq \rho(a^\tau)$ implying that $t_{\tau+2}^L \geq t_\tau^L$. Iterating the argument gives that $t_{\tau'}^L \geq t_\tau^L$ for all $\tau' > \tau$. Thus, t_τ^L is non-decreasing in τ. Note that because $-1 \leq t_\tau^L \leq x^*$ the $\lim_{\tau\to\infty} t_\tau^L$ exists. Similarly the $\lim_{\tau\to\infty} t_\tau^H$ exists. Now, by Theorem 1, $\rho(a^\tau) \geq G(\rho(a^{\tau-1}), x^*)$. Thus, if $\rho\left(a^{\tau-1}\right) \leq \overline{\gamma}$ then $\rho\left(a^\tau\right) \geq G(\rho(a^{\tau-1}), x^*) > \rho\left(a^{\tau-1}\right)$. Suppose that $\lim_{\tau\to\infty} \rho(a^\tau) = \rho^* \leq \overline{\gamma}$. Let $t_*^- = x^*(1-\rho^*)-\rho^*$ and $t_*^+ = x^*(1-\rho^*)+\rho^*$. We know that $G(\rho^*, x^*) > \rho^*$ but, this implies, for sufficiently large τ, that $G(\rho(a^\tau), x^*) > \rho^*$as desired. ■

Proof of Corollaries 1 and 2: We make use of the concepts of stochastic stability and a regular perturbation of a Markov Process derived from the work of Freidlin and Wentzell (1984) and developed by, amongst others, Young (1993). We provide here a very informal discussion of the issues and the reader is advised to consult Young (1993) for a more complete discussion. The analysis assumes a non-ergodic Markov Process P^0. This process is perturbed by shocks that occur with probability λ to give an ergodic Markov Process P^λ. Process P^λ has a unique stationary distribution θ^λ where θ_a^λ denotes the cumulative relative frequency of state a. A state a is *stochastically stable* if $\lim_{\lambda\to 0} \theta_a^\lambda > 0$. Thus, for small λ, only stochastically stable states are observed with a probability significantly different to zero. The process P^0 gives rise to a set of stationary states Σ^S. When $\lambda > 0$ transition from $a \in \Sigma^S$ to $a' \in \Sigma^S$ is possible but requires shocks. The *stochastic potential* of a stationary state a is the sum, over all stationary states $a' \in \Sigma^S$, of the minimum number of shocks the transition from a' to a would require. The stochastically stable states (Theorem 4 of Young 1993) are those with minimum stochastic potential. That is they are the stationary states that

can be reached with fewest shocks.

Now consider the dynamic of this paper equating states with action profiles. For each action $x \in X$ we can derive, assuming that $\lambda = 0$, the action profile $\overrightarrow{a}^{\tau}_{x}$ that occurs in period τ given initial state a^0_x where $\mu(a^0_x) = x$ and $\rho(a^0_x) = \varepsilon$. Let $NC \subset X$ denote the set of actions where $x \in NC$ if and only if there exists a finite τ such that $\overrightarrow{a}^{\tau}_{x} = \overline{a}$ (where $\overline{a}$ is the non-conformist equilibrium). If $x \in NC$ then no shocks would be required for x to be replaced as a norm. Let $\rho^*_x := \lim_{\tau\to\infty} \overrightarrow{a}^{\tau}_{x}$. Given a real number ρ let $s(\rho) := \{\min y \in \mathbb{Z} : \rho - y\varepsilon \leq \varepsilon\}$. Informally, if $x \notin NC$ is the norm then $s(\rho^*_x)$ is the number of shocks that would need to occur for x to no longer be the norm. For each $x \notin NC$ set $d(x) := \left\{\min d \geq 1 : s(\overrightarrow{a}^{d}_{x}) = s(\rho^*_x)\right\}$.[23] A finite $d(x)$ exists. Note that if $\rho^*_x \leq \varepsilon$ but $x \notin NC$ then $d(x) = 1$. For each action $x \notin NC$ let $\overrightarrow{a}_x := \overrightarrow{a}^{d(x)}_{x}$ and let $s_x := s(\overrightarrow{a}_x)$. For each $x \in NC$ set $s_x = 0$. As we discuss in the final paragraph, $\overrightarrow{a}_x$ can be seen as a 'representative' of the action profile that will occur in the long run if x is the norm and there are no shocks; integer s_x is the number of shocks that would be required for x to be replaced as a norm. [To avoid technical complications we assume that $\rho(\overrightarrow{a}_x)$ is not divisible by ε. See the end of the proof for further comment].

One complication of using stochastic stability is that we require a finite state space. We 'construct one' by amending the 'original' dynamic. For each $x \notin NC$ and any positive integer $b \leq s_x$ let $\overrightarrow{a}^{-b}_{x}$ denote a strategy profile where $\mu(\overrightarrow{a}^{-b}_{x}) = x$ and $\rho(\overrightarrow{a}^{-b}_{x}) = \rho(\overrightarrow{a}_x) - b\varepsilon$. We can think of $\overrightarrow{a}^{-b}_{x}$ as the action profile that results if b consecutive shocks occur from action profile $\overrightarrow{a}_x$. In the amended dynamic when $\lambda = 0$ we assume for all $x \notin NC$ that (1) $\overrightarrow{a}_x$ is a stationary state and (2) action profile $\overrightarrow{a}_x$ occurs in period $\tau+1$ if action profile a^0_x or $\overrightarrow{a}^{-b}_{x}$, for any b, occur in period τ. For all $x \in NC$ we assume that (3) action profile $\overline{a}$ occurs in period $\tau + 1$ if action profile

[23] If there exists no such d then put $d(x) = \left\{\min d \geq 1 : s(\overrightarrow{a}^{d}_{x}) = s(\rho^*_x) - 1\right\}$. This can occur if $\lim_{\tau\to\infty} \overrightarrow{a}^{\tau}_{x}$ is a multiple of ε but sequence $\{\overrightarrow{a}^{\tau}_{x}\}$ never attains its limit.

a_x^0 occurs in period τ. Essentially, this condenses the 'original' dynamic to a summarized version where, if x is the norm, we 'fast forward' the dynamics to go immediately to either state $\overrightarrow{a}_x$ or $\overline{a}$ and then we 'truncate' the dynamic by assuming that nothing else would change. Partition action space X into subsets $NC, X_1, .., X_\varsigma$ such that $x \in X_y$ if $s_x = y$. Let X^F denote a finite subset of X that contains at least one action $x \in X_y$ or NC for every subset X_y and NC that has positive measure (according to distribution F). Set X^F is representative of the number of shocks required to move between norms. The final step in defining the amended dynamic is to assume that shocks can only occur at actions $x \in X^F$ and that if $\mu(a^\tau) = x^*$ then a shock cannot occur at x^*.

Given this amended dynamic we can consider finite state space Σ that contains a_x^0 for all $x \in X^F$, $\overline{a}$ if NC is non-empty and $\overrightarrow{a}_x$, $\overrightarrow{a}_x^{-1}$, ..., $\overrightarrow{a}_x^{-s_x}$ if $x \in X^F \backslash NC$. If $\lambda = 0$ then the amended dynamic is a deterministic Markov chain on state space Σ. Denote the transition matrix by P^0. Let Σ^S denote the set of stationary states where $\overrightarrow{a}_x \in \Sigma^S$ for all $x \in X^F \backslash NC$ and $\overline{a} \in \Sigma^S$ if NC is non-empty. If $\lambda > 0$ then we have an aperiodic and irreducible Markov Process on state space Σ with transition matrix P^λ. The family of Markov Processes P^λ is a regular perturbation of P^0 as defined by Young (1993).

The transition from $\overrightarrow{a}_x$ to any other $a \in \Sigma^S$ requires s_x shocks. The transition from $\overline{a}$ to $\overrightarrow{a}_x \in \Sigma^S$ requires one shock. Thus, applying Theorem 4 of Young (1993), if action profile $\overrightarrow{a}_{x^*}$ is stochastically stable then $s_{x^*} = \max_{x \in X^F} s_x$. If there exists $\overrightarrow{a}_x$ such that $s_x \geq 2$ then $\overline{a}$ is not stochastically stable. We now prove Corollaries 1 and 2 given the amended dynamic.

Corollary 1: if f is not uniform then there exists some $\mu \in X$ and $\nu > 0$ such that $f(x) > \frac{1}{2}$ for all $x \in (\mu - \nu, \mu + \nu)$. Thus, there exists $\overline{\nu}, \overline{\gamma} > 0$ such that $G(\gamma, x^*) > \gamma$ for all $x^* \in (\mu - \overline{\nu}, \mu + \overline{\nu})$ and $\gamma \in (0, \overline{\gamma})$. Applying Lemma 6, for sufficiently small ε, $\rho_x^* > 2\varepsilon$ for all $x \in (\mu - \overline{\nu}, \mu + \overline{\nu})$. Thus there exists $x \in X_y$ for some $y \geq 2$. This gives the desired result. If f is uniform then

$G(\gamma, x^*) = \gamma$ and so $\rho_x^* \geq \varepsilon$ for all x. This means that NC is the empty set and Corollary 1 is immediate.

Corollary 2: if there exists $x^* \in X$ such that $G(\gamma, x^*) > \gamma$ for all γ then by Lemma 6 and continuity of f there exists region $(x_1, x_2) \ni x^*$ such that $\rho_x^* > 1 - \varepsilon$ for all $x \in (x_1, x_2)$. Clearly, s_x achieves its maximum for such x proving Corollary 2.

Finally we can comment on why outcomes given the amended dynamic are consistent with those of the original dynamic. Lemma 6 and the properties of G allow us to truncate the dynamic at $\overrightarrow{a}_x$ and fast forward to go immediately to $\overrightarrow{a}_x$. This is because we know, from Lemma 6, that for any state a_x^τ where $\mu(a_x^\tau) = x$ and $\rho(a_x^\tau) < \rho_x^*$ the dynamic will evolve, in the absence of future shocks, until the number of shocks required to 'escape an x norm' is given by $s_x = s(\overrightarrow{a}_x)$. Thus, we are not altering the likelihood of a transition between norms. Given the nature of X^F, and for the properties of the dynamic of interest here, assuming that shocks can happen at only a finite set of actions is also acceptable. This is because X^F includes a 'representative' from each X_y and so we are capturing the likelihood of 'escaping' from norm x to some other norm $x' \neq x$. Also, stochastic stability does not depend on the relative probabilities of a shock occurring at each state (Young 1993). Thus, we can take just one representative from each X_y. That positive shocks could occur twice for the same action (possible in the finite setting) is directly ruled out. Finally, we are imposing that the initial state belong to Σ. Given, however, that consecutive shocks can erode any norm and our focus is on long run dynamics there is no loss in generality in assuming the initial state is $\overline{a}$ or a_x^0 for some x.

[If $\rho(\overrightarrow{a}_x)$ is divisible by ε then we have to do slightly more. This is because if s_x shocks occur then $\rho(x, a) = \varepsilon$ and $\rho(x', a) = \varepsilon$ for some $x' \neq x$ and so there is no norm. This means that $\overline{a}$ must belong to Σ. It also means the transition from $\overrightarrow{a}_x$ to $\overrightarrow{a}_{x'}$ requires $s_x + 1$ shocks. Note, however, that

the number of shocks required to reach $\overline{a}$ does not change. Further, our proof does not use that $\overline{a} \notin \Sigma$. Thus, the conclusions remain valid.]■

Proof of examples of distributions: We consider the three examples in turn. First, $f(t) = 1 - |t|$. Then

$$F(t) = \begin{cases} \frac{1}{2} + t + \frac{t^2}{2} \text{ if } t \in [-1, 0] \\ \frac{1}{2} + t - \frac{t^2}{2} \text{ if } t \in (0, 1]. \end{cases}$$

Fix a $\mu \leq 0$ and a γ. Suppose that $\mu - \gamma\mu + \gamma \leq 0$. Then

$$\begin{aligned} G(\gamma) &= (\mu - \gamma\mu + \gamma) + \frac{1}{2}(\mu - \gamma\mu + \gamma)^2 - (\mu - \gamma\mu - \gamma) - \frac{1}{2}(\mu - \gamma\mu - \gamma)^2 \\ &= 2\gamma + 2\gamma\mu - 2\gamma^2\mu. \end{aligned}$$

Thus $G(\gamma) > \gamma$ if $\gamma(1+2\mu) > 2\gamma^2\mu$. We know that $\gamma^2\mu \leq 0$ and $\gamma(1+2\mu) > 0$ if $\mu > -\frac{1}{2}$ giving the desired result. If $\mu - \gamma\mu + \gamma > 0$ then

$$\begin{aligned} G(\gamma) &= (\mu - \gamma\mu + \gamma) - \frac{1}{2}(\mu - \gamma\mu + \gamma)^2 - (\mu - \gamma\mu - \gamma) - \frac{1}{2}(\mu - \gamma\mu - \gamma)^2 \\ &= 2\gamma - \mu^2 + 2\gamma\mu^2 - \gamma^2\mu^2 - \gamma^2. \end{aligned}$$

Thus $G(\gamma) > \gamma$ if $\gamma > \mu^2(1 - \gamma)$. We know that $\mu - \gamma\mu + \gamma > 0$ so $\gamma > -\mu(1 - \gamma)$. Thus we have the desired result if $0 \geq \mu \geq -1$. A symmetric argument treats $\mu > 0$.

Now, let $f(t) = \frac{1}{2} - \frac{1}{2}t$. This implies that

$$F(t) = \frac{3}{4} + \frac{t}{2} - \frac{t^2}{4}$$

for all t. Thus,

$$\begin{aligned} G(\gamma) &= \frac{1}{2}(\mu - \gamma\mu + \gamma) - \frac{1}{4}(\mu - \gamma\mu + \gamma)^2 - \frac{1}{2}(\mu - \gamma\mu - \gamma) + \frac{1}{4}(\mu - \gamma\mu - \gamma)^2 \\ &= \gamma - \frac{1}{2}\gamma\mu + \frac{1}{2}\gamma^2\mu. \end{aligned}$$

So, $G(\gamma) > \gamma$ if $\frac{1}{2}\gamma\mu > \frac{1}{2}\mu$ or $\mu < 0$.

Finally, let f be symmetric around 0 with $f(t) = \frac{1}{4}$ for $t \in [0, 0.1)$; $f(t) = 4$ for $t \in [0.1, 0.2]$; $f(t) = \frac{15}{64}(1-t)$ for $t \in (0.2, 1]$. The full proof is somewhat tedious but we can sketch the proof for the seemingly most unlikely case of $\mu = 0.1$. When $\gamma \leq \frac{1}{9}$ we have $\mu - \gamma\mu + \gamma \leq 0.2$. So, $G(\gamma) \geq F(0.1 - 0.1\gamma + \gamma) - F(0.1) \geq 4[0.1 - 0.1\gamma + \gamma - 0.1] > \gamma$. For $\frac{1}{9} \leq \gamma \leq \frac{3}{11}$ we have $G(\gamma) \geq F(0.2) - F(0.1) = \frac{2}{5} > \gamma$. Note that $\mu - \gamma\mu - \gamma = -0.2$ when $\gamma = \frac{3}{11}$. So, for $\gamma \geq \frac{3}{11}$ we have $G(\gamma) \geq F(0.2) - F(-0.2) = \frac{17}{20}$. In summary, we have shown that $G(\gamma) > \gamma$ for all $0 < \gamma < \frac{17}{20}$. The remaining γ can be treated as in the first example ($f(t) = 1 - |t|$).■

Proof of Corollary 5: Applying Lemmas 3 and 4 we know that if Properties 1 and 2 hold, conformity is an equilibrium and $\rho(\mu, l, a) = \varrho$ then $\mu = \mathcal{BR}_t(a)$ for all agents at location l with type $t \in [t^-_\varrho, t^+_\varrho]$ where $t^+_\varrho := (1-\varrho)\mu + \varrho$ and $t^-_\varrho := (1-\varrho)\mu - \varrho$. If Property 1 holds with equality, Property 2 holds and conformist equilibrium a^μ is not strict then $\mu = \mathcal{BR}_t(a)$ only for those agents at location l with type $t \in [t^-_\varrho, t^+_\varrho]$. The statement of the Corollary is now trivial.■

References

[1] Akerlof, G. A., (1980) 'A Theory of Social Custom of which Unemployment May Be One Consequence', *Quarterly Journal of Economics* 94: 749-75.

[2] Azar, O.H., (2004) 'What sustains social norms and how they evolve? The case of tipping', *Journal of Economic Behavior and Organization* 54: 49-64.

[3] Bernheim, B. D., (1994) 'A Theory of Conformity', *Journal of Political Economy* 102: 841-877.

[4] Bikhandani, S., D. Hirshleifer and I. Welch, (1992) 'A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades', *Journal of Political Economy* 100: 992-1026.

[5] Elster, J., (1989) 'Social Norms and Economic Theory', *Journal of Economic Perspectives* 3: 99-117.

[6] Freidlin, M. and A. Wentzell, (1984) *Random Perturbations of Dynamical Systems* New York: Springer Verlag.

[7] Fudenberg, D. and D.K. Levine, (1998) *The Theory of Learning in Games*, MIT Press.

[8] Jones, S. R. G., (1984) *The Economics of Conformism* Oxford. Blackwell.

[9] Juang, W., (2001) 'Learning from Popularity', *Econometrica* 69: 735-747.

[10] Kreps, D., (1997) 'Intrinsic motivation and extrinsic incentives', *The American Economic Review* 87: 359-364.

[11] Lewis, D., (1967) *Convention: A Philisophical Study*, Cambridge, Mass: Harvard University Press.

[12] Lindbeck, A. (1997) 'Incentives and Social Norms in Household Behavior', *The American Economic Review* 87: 370-377.

[13] Lindbeck, A., S. Nyberg and J. Weibull (1999) 'Social norms and Economic Incentives in the Welfare State' *The Quarterly Journal of Economics* 114: 1-35.

[14] Morris, S., (2000) 'Contagion', *Review of Economic Studies* 67: 57-78.

[15] Ochs, J., and I.-U. Park (2004) 'Overcoming the Coordination Problem: Dynamic Formation of Networks', Working Paper.

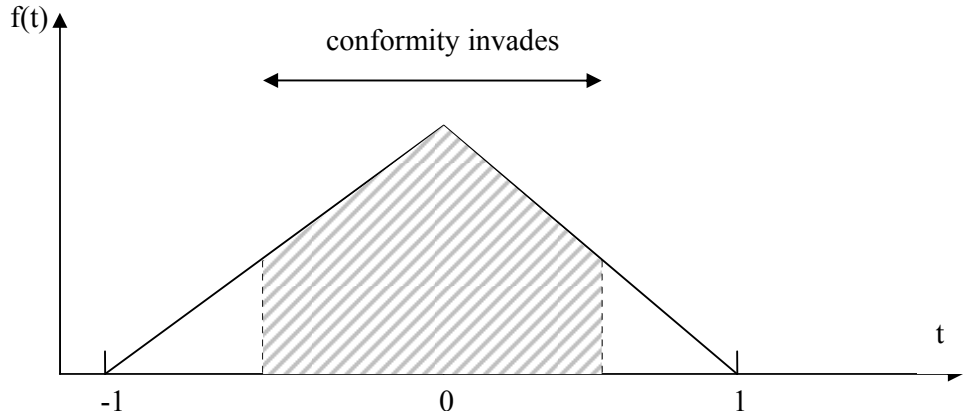
[16] Weibull, J. (1996) *Evolutionary Game Theory* MIT Press, Cambridge Massachusetts.

[17] Wooders, M., E. Cartwright and R. Selten (2006) "Behavioral conformity in games with many players," *Games and Economic Behavior* 57: 347-360.

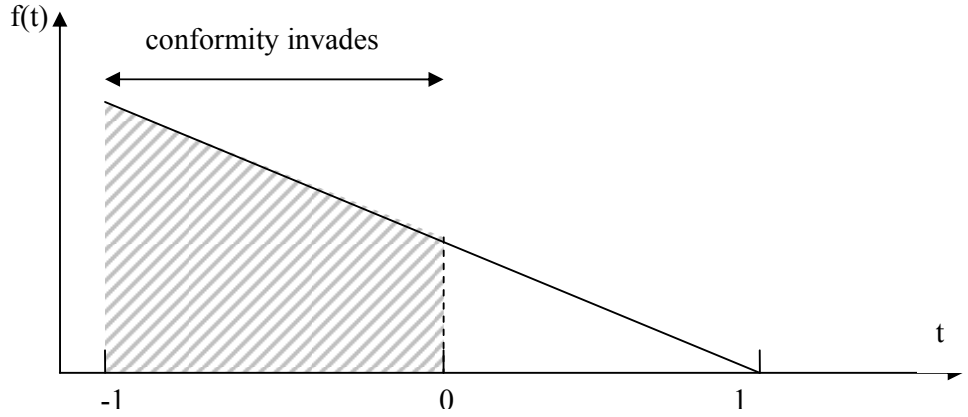
[18] Young, P., (1993) 'Evolution of conventions', *Econometrica* 61: 57-84.

Figure 1: The range of μ for which conformity can invade in different populations:

1a: $f(t) = 1 - |t|$.



1b: $f(t) = 0.5(1 - t)$.



1c: f is bimodal.

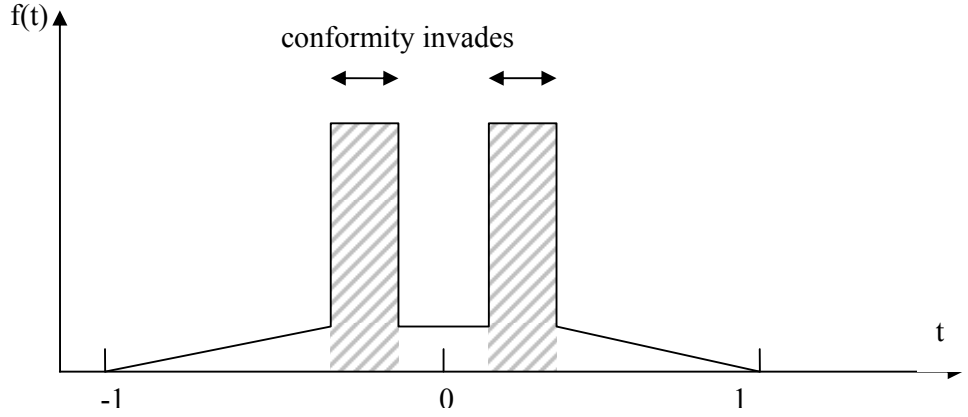
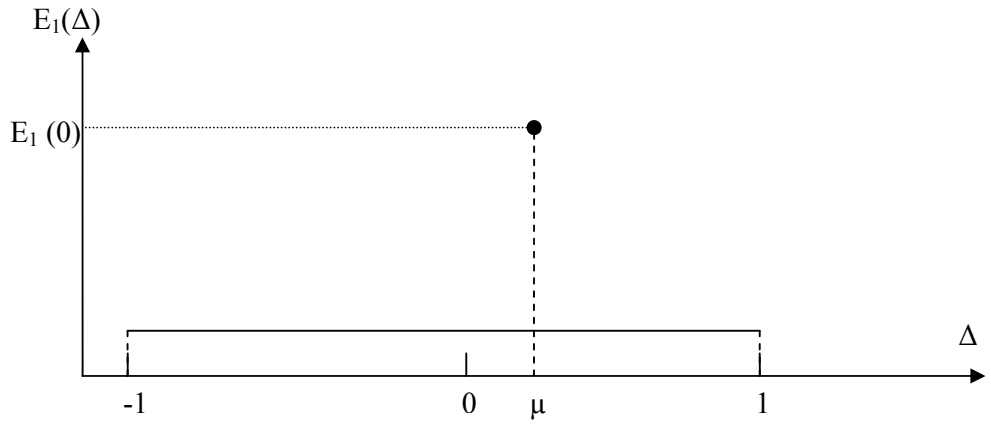
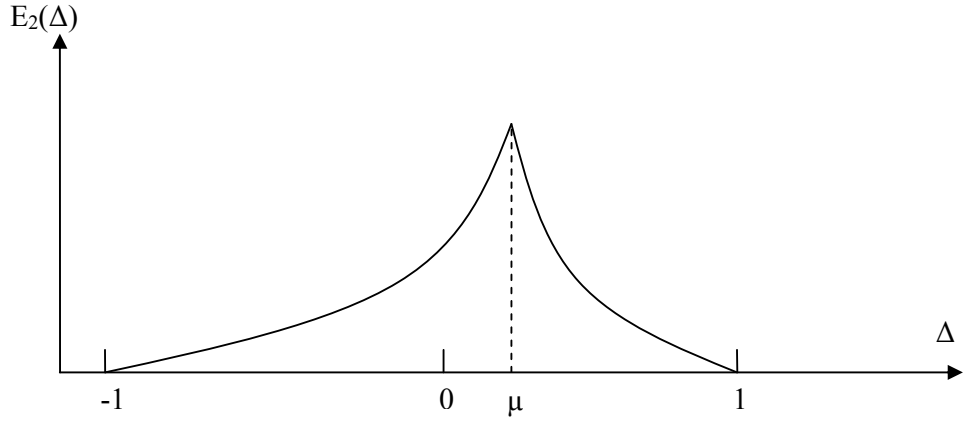


Figure 2: Examples of esteem functions.

2a: Discontinuous and non-discriminatory.



2b: Convex and continuous.



2c: Concave and discontinuous

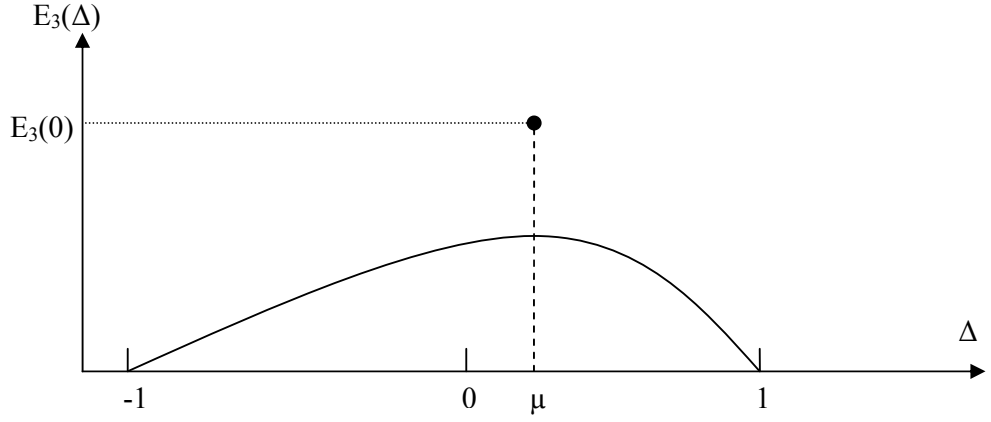
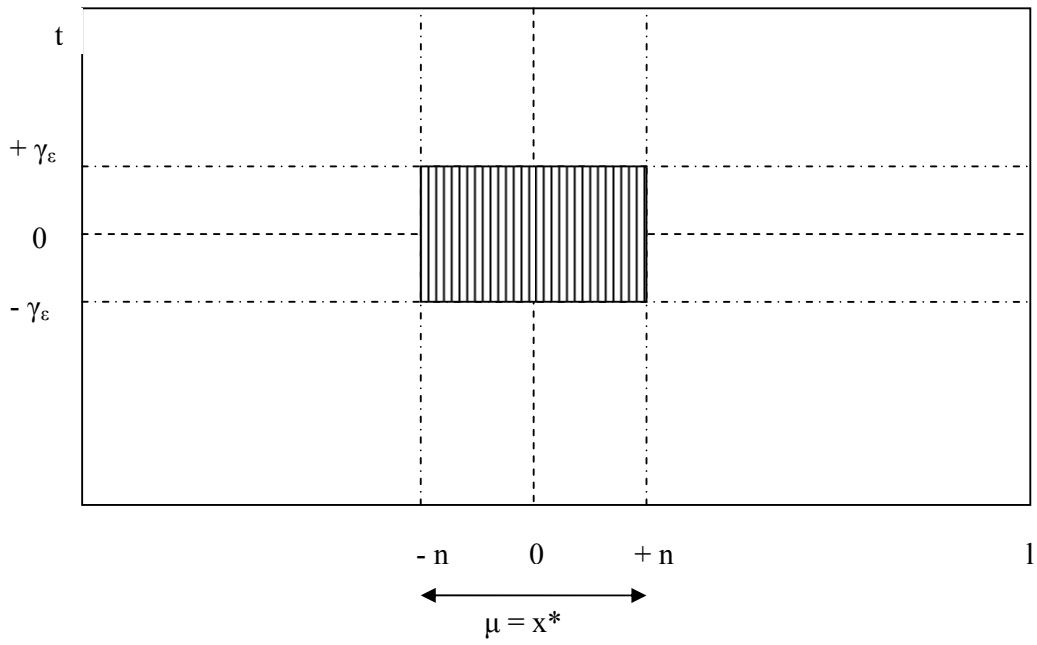


Figure 3a: Type and location combinations for agents who conform for Example 5.

Period 2.



Period 3.

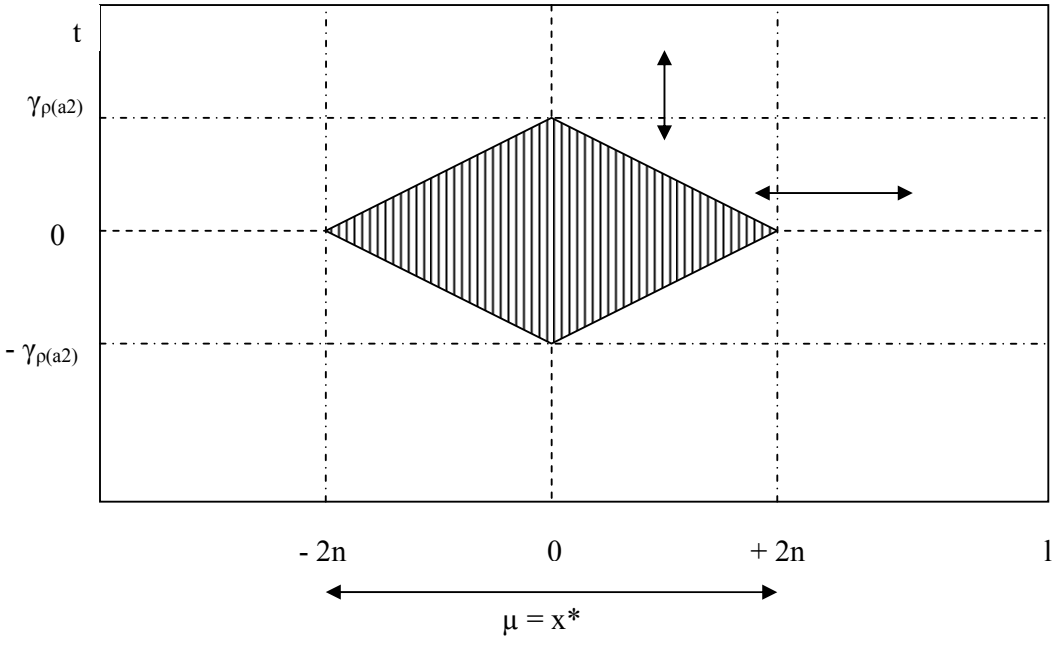
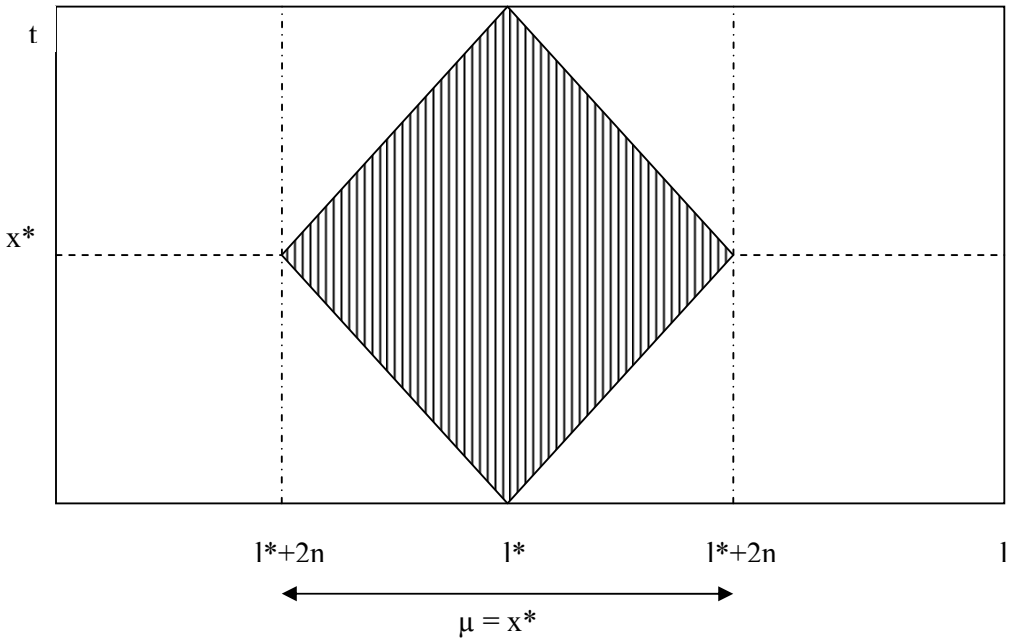


Figure 3b: Type and location combinations for agents who conform for Example 6.

Period 3.



Period 4

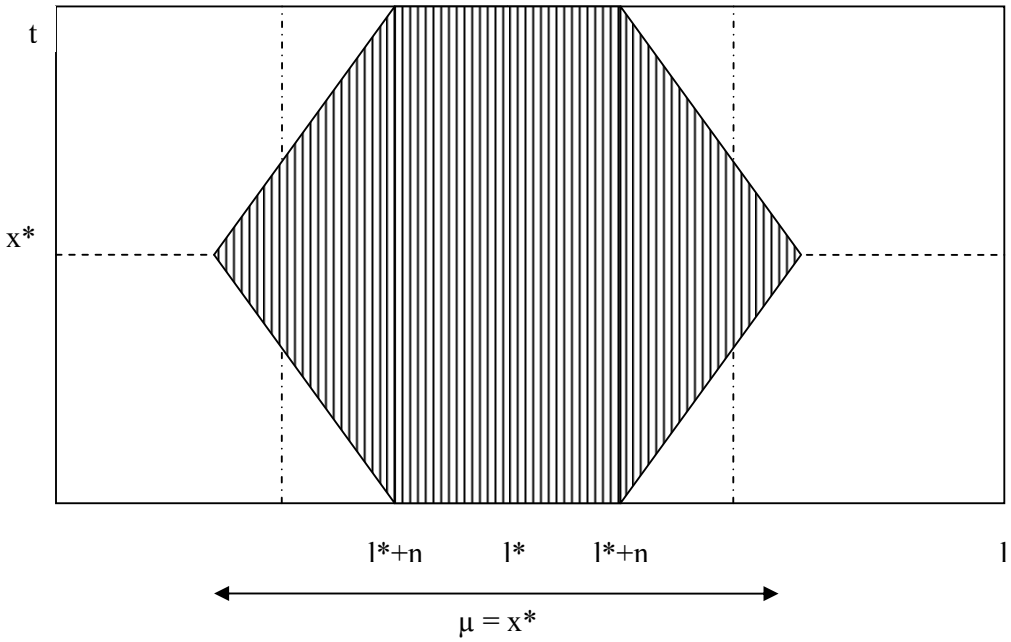


Figure 4: The coexistence of norms: