A Note on Endogenous Norms in A Theory of Conformity

Ben Gillen[*]
California Institute of Technology

September 12, 2014

Abstract

This paper analyzes conformist tendencies for a population in which individuals gain utility by mimicking the average behavior, characterizing norms by the mean behavior. In so doing, the model extends Bernheim (1994)'s "A Theory of Conformity" by introducing an endogenous mechanism for establishing social norms. The most interesting result is that this extension does not alter the properties of equilibria established in Bernheim's initial development, that is, social preferences generally give rise to more concentrated behavior and a conformist pool forms when social preferences are sufficiently prominent. Further the extension introduces no new equilibria, since even though Bernheim's development included a multiplicity of locations for conformist outcomes, these outcomes are identified exactly by the location of the social norm within the extended model. In addition to illustrating the determinants of conformist behavior with an endogenous reference point, these findings support applied work inferring social norms from average behavior.

[*] California Institute of Technology, Division of Humanities & Social Sciences, 1200 E California Blvd, MC 228-77, Pasadena, CA 91125. This paper is taken from my doctoral thesis at the University of California, San Diego. I started this project with a lot of help from Navin Kartik and have benefited from the comments of S. Nageeb Ali, B. Douglas Bernheim, Vincent Crawford, David Miller, Allan Timmermann, and seminar participants at UCSD. I'm particularly grateful for the Research Assistance of Khai Chiong, who helped correct an error in previous drafts. All remaining mistakes are my own. Comments, feedback, and suggestions are welcome at bgillen@caltech.edu.

Keywords: Applied Microeconomics, Conformity, Social Preferences, Signaling Models
JEL Codes: C7, D7, D8

1 Introduction

Convergent and conformist behavior arises in a wide variety of economic contexts. An extensive literature analyzing herd behavior has developed in the context of coordination games. The outcome of convergent behavior in these models is not surprising, given mutually reinforcing preferences implemented by construction. This coordination does not capture the essence of conformity, which corresponds to a spontaneous coordination across individuals despite heterogeneous preferences. As such, a model of conformity requires a formulation of social preferences that induce coordination despite varied private preferences.

This note extends the completely continuous, preference-based approach to conformity initially developed in Bernheim (1994)'s "A Theory of Conformity," embedding an endogenous model for norm formation. Bernheim (1994) analyzes a signaling game where individuals balance their privately intrinsic preferences with a desire to be perceived as the socially ideal type. Andreoni and Bernheim (2009) apply the model to analyze experimental subjects' behavior in light of social norms in dictator games.[1] Here, the normative behavior is endogenously determined based on the chosen actions of all agents, defining the social norm as the population expected action (though they can be readily extended to alternative mechanisms establishing the social norm). A player's social utility is derived from being perceived as the type that truly desires to take the average action selected by the population rather than someone that is simply shading their behavior in accordance with social norms. Intuitively, these incentives reflect an individual's desire to be perceived as a trend-setter (a "true" fan of the fad) rather than a trend-follower (a "faker" following others' lead).

[1] This formulation captures settings where conformity arises from peer effects, as in Akerlof (1980), or due to post-game according of social of status. Cole, Mailath, and Postlewaite (1992) lay foundations for alternatively characterizing social preferences as peer-group rank. Benabou and Tirole (2006) embed monotonic social preferences in a signaling framework that trades off social utility against a desire to appear disinterested. A very different sort of spontaneous conformist behavior arises in the presence of incentives to strategically report information, as in the information cascades of Banerjee (1992) and Bikhchandani, Hirschleifer, and Welch (1992).

2 The Model for Preferences and Actions

The model follows Bernheim (1994) with a large number, I, of individuals, indexed by i, who are each privately assigned a type $t_i \in [0, 2] \equiv T$. Players' types are privately observed, but each player chooses a publicly observable action, $a_i \in [0, 2] \equiv A$, that may depend on their true type. The types are drawn independently from T according to the distribution $F(\cdot)$, with continuous density $f(\cdot)$ bounded away from zero, and $F(2) = 1$.

Preferences reflect an individual's *intrinsic* and *social* utility. The individual's type (t) represents their "Intrinsic Bliss Point" (IBP). Intrinsic utility rewards actions close to an invidiual's IBP according to the function $g(a - t)$. An individual's social utility is maximized when their perceived type, based on their action, is near a Social Bliss Point" (SBP) denoted by α. Letting b_i representing the agent's perceived type, these preferences are reflected in the function $h(b_i - \alpha)$. Both g and h are maximized at zero, twice continuously differentiable, strictly concave, and symmetric, mainly to ensure the conformity result is not driven by a discontinuity in preferences.

With social bliss point α, a player's total utility given their type t, action a, and perceived type b, combines intrinsic and social utility:

$$u(a, t; b, \alpha, \lambda) = g(a - t) + \lambda h(b - \alpha) \tag{1}$$

The weight on an agent's social utility, λ, is referred to as the *social preference intensity.*

To link the social bliss point to players' behavior, suppose the SBP matches the expected action, i.e., $\alpha = E_f[a(t)]$.[2] An inference function $\phi(b, a; \alpha, \lambda)$ represents the probability a player assigns to being perceived as type b when taking the action a. As the SBP is influenced by others' behavior, players also form beliefs about α, here represented by the distribution

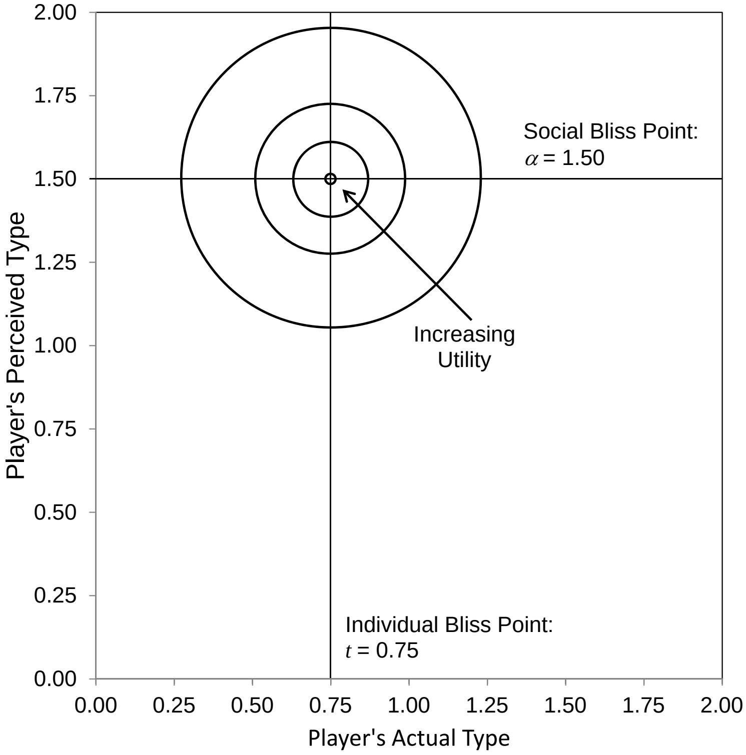
[2] With many players, the SBP could be a measurable function of players' observed actions, such as the average action actually chosen by players in the game. The law of large numbers ensures the basic equilibrium results go through directly, with the exception of (Theorem 3), discussed below.

$\pi\left(\alpha;\lambda\right)$. These beliefs define an individual's utility maximization problem:

$$\begin{aligned}\max_{a\in A} E\left[u\left(a,t;\alpha,\lambda\right)\right] \qquad (2)\\ = g\left(a-t\right)+\lambda\int_{\hat{\alpha}\in T}\left(\int_{b\in T} h\left(b-\hat{\alpha}\right)\phi\left(b,a;\hat{\alpha},\lambda\right)db\right)d\pi\left(\hat{\alpha};\lambda\right)\end{aligned}$$

With the social bliss point matching the population expected action, the beliefs $\pi\left(\alpha;\lambda\right)$ reduce to a degenerate point distribution and the double integral becomes a single expectation. Letting $\hat{\alpha}=E_{\pi}\left[\alpha\right]$, the optimization problem becomes:

$$\max_{a\in A} E\left[u\left(a,t;\hat{\alpha},\lambda\right)\right]=g\left(a-t\right)+\lambda\int_{b\in T} h\left(b-\hat{\alpha}\right)\phi\left(b,a;\hat{\alpha},\lambda\right)db \qquad (3)$$



The Bernheim (1994) "spherical case" sets $g\left(z\right)=-z^2$ and $h\left(b;\alpha\right)=-\left(b-\alpha\right)^2$. An agent's indifference curves in the $\left(a,b\right)$ plane appear as concentric circles centered on the point $\left(t,\alpha\right)$.

Figure 1: Agent Preferences in the "Spherical Case"

Figure 1 illustrates agent preferences in a quadratic example. Generally, the indifference curves are horizontal at $a=t$ and symmetric around the line $a=t$, while vertical at $b=\hat{\alpha}$

and symmetric around the line $b = \hat{\alpha}$. Note that indifference curves for players of different type do not satisfy single-crossing property.

3 Characterizing Equilibrium

This section characterizes the equilibria in the game, specifying the following equilibrium conditions for an interim Bayes Perfect Nash Equilibrium:

1. An action function, $a^* (t; \alpha, \lambda, \phi) : T \to A$, such that for all $a' \in A$ and $t \in T$,

$$U (a^* (t; \alpha, \lambda, \phi), t; \alpha, \lambda, \phi) \geq U (a', t; \alpha, \lambda, \phi)$$

2. A conditional inference function, $\phi (b, a; \alpha, \lambda)$, representing a probability distribution over the agent's inferred type, b, given their action a. The inference function must be consistent with Bayes' Rule along the equilibrium path.

3. Beliefs about the expected average action, denoted by the probability distribution π with mean $E_\pi [\alpha|\lambda] = \alpha^* (\lambda) = \int_T a^* (t; \alpha^* (\lambda), \lambda) f (t) \, dt$ and variance going to zero as I gets large.

With an exogenously fixed social bliss point, a Bayes Perfect Nash Equilibrium only requires satisfying conditions (1) and (2). Define an action function and conditional inference function satisfying these conditions for a fixed social bliss point to be an *exogenous social equilibrium*. In any exogenous social equilibrium, Bernheim (1994)'s result that players' optimal actions are monotonic in their types applies with no modification to the original proof.

Bernheim Theorem 1 *If $t > t'$, then the utility-maximizing action function in any equilibrium must satisfy $a^* (t; \alpha, \lambda, \phi) \geq a^* (t'; \alpha, \lambda, \phi)$.*

3.1 Fully Separating Equilibrium

The model supports full separation with each type choosing a different equilibrium action if and only if the social preference intensity is sufficiently small. Extending this result from Bernheim (1994) to the present analysis focuses on condition 3, which leads players to behave as if the social bliss point were exogenously fixed by the equilibrium.

The fully separating exogenous social equilibrium for a fixed α is identified by an inference characterizing function $\phi_s(a)$ that assigns probability one to a player being accorded status consistent with their true type.

$$\phi(b,a) = \begin{cases} 1 & \text{if } b = \phi_s(a) \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Differentiating the utility from equation 1, the slope of indifference curves in the (a,b) plane for type t players is:

$$\frac{db}{da} = -\frac{g'(a-t)}{\lambda h'(b-\alpha)} \tag{5}$$

This slope reflects a player's marginal rate of substitution between internal consistency and social edification. The relative price an agent faces in this trade-off is the slope of the inference characterizing function. Utility is maximized when agents' indifference curves are tangent to the inference characterizing function.

$$\phi_s'(a;\alpha) = -\frac{g'(a-t)}{\lambda h'(\phi_s(a;\alpha)-\alpha)} \tag{6}$$

In equilibrium, the inference function must truly reveal individuals' types, substituting t out of the condition:

$$\phi_s'(a;\alpha) = -\frac{g'(a-\phi_s(a;\alpha))}{\lambda h'(\phi_s(a;\alpha)-\alpha)} \tag{7}$$

Equation 7 simply introduces α to Bernheim (1994)'s equation (12) that defines a differential equation identifying ϕ_s. The extreme types $\{0,2\}$ have no incentive to deviate from actions matching their IBP, giving initial conditions:

$$\phi_s(0;\alpha) = 0, \text{ and, } \phi_s(2;\alpha) = 2 \tag{8}$$

For fixed α, the analysis here is no different from Bernheim (1994). Equation 7, along with initial conditions 8, defines two differential equations in the (a, b) plane. The first ODE starts from $(0, 0)$ and moves northeast while the second starts from $(2, 2)$ and moves southwest. A fully separating equilibrium requires the paths for these two ODE's to meet, crossing the 45-degree line at the SBP.[3] Figure 2 illustrates the result when $\lambda \leq 0.25$.

An exogenous social equilibrium must also satisfy equilibrium condition 3 to be an endogenous social equilibrium. Lemma 1 establishes that, over all the exogenous social equilibria, there is only one such endogenous social equilibrium.

Lemma 1 *Given an optimal action function $a^* (t; \alpha, \lambda)$ that is strictly monotonic in t and a consistent inference function $\phi (b, a; \alpha, \lambda)$ for every $\alpha \in T$, there exists a unique α^* such that $\alpha^* = E_f [a^* (t; \alpha^*, \lambda, \phi)]$.*

The crux of the proof to Lemma 1 in Appendix A1 requires showing that the mapping $\gamma_\lambda (\alpha) : \alpha \mapsto \int_T a^* (t; \alpha, \lambda) f (t) \, dt$ is a contraction mapping for any λ. This contraction result arises from the intuition that the average player's behavioral response to a change in the SBP is strictly less than the change in the SBP itself due to the moderating effect of intrinsic preferences.

Fixing α, there is a critical social preference intensity $\lambda^* (\alpha)$ such that a fully separating exogenous social equilibrium exists if and only if $\lambda \leq \lambda^* (\alpha)$. This result is stated exactly as in Bernheim (1994)'s Theorem 2 and proved in the appendix following Bernheim's reasoning closely, with minor technical development.

Bernheim Theorem 2 *For each α, there exists $\lambda^* (\alpha) > 0$ such that a fully separating equilibrium exists if and only if $\lambda \leq \lambda^* (\alpha)$.*

[3] Bernheim's analysis applies with only minor adjustments to notation. Consider the range of the inference characterizing function for agents with $t \leq \alpha$. Define $\underline{A} = \phi_s^{-1} ([0, \alpha])$ as the choices by individuals with $t \in [0, \alpha]$. By monotonicity, $\underline{A}$ is an interval $[0, \underline{a}]$ over which standard arguments establish existence and uniqueness of ϕ_s. Over $\underline{A}$, $\phi_s (a) \leq a$, yielding the implication $\underline{a} \geq \alpha$. Now consider the complement of the type space, agents with $t \geq \alpha$. Defining $[\bar{a}, 2] = \bar{X} = \phi_s^{-1} ([\alpha, 2])$, parallel arguments imply $\bar{a} \leq \alpha$. As a fully separating equilibrium requires that a unique type be assigned to each action, its existence requires $\underline{a} = \bar{a} = \alpha$.

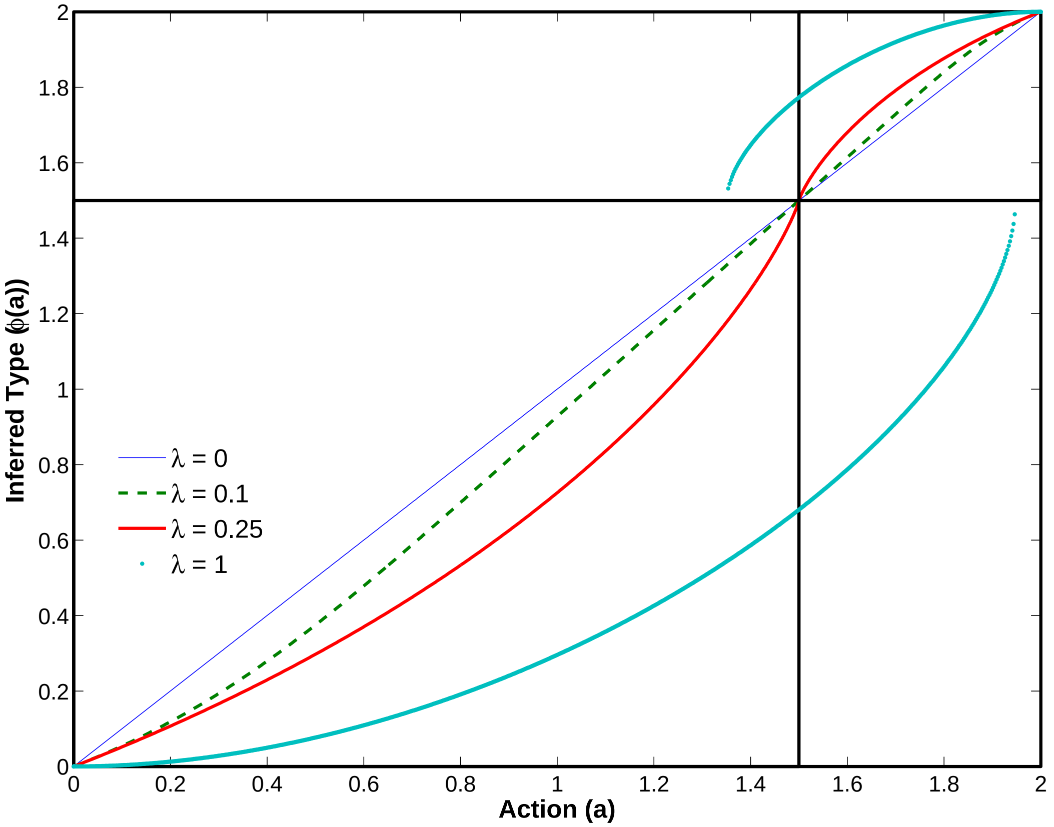


Figure 2: Separating Inference Function in Spherical Case when $\alpha = 1.5$

Figure 2 illustrates the ODE's defining the inference characterizing function for the example with quadratic preferences from Figure 1.

Surprisingly, even though the critical social preference intensities are specifically identified for each α, these thresholds are insensitive to the SBP's location. As such, $\lambda^* (\alpha) = \lambda^* (\alpha')$, establishing the following Theorem:

Theorem 1 *There exists a unique* $\lambda^* > 0$ *such that a fully separating equilibrium exists if and only if* $\lambda \leq \lambda^*$

3.2 Equilibria with Incomplete Separation

Rich signaling games commonly generate a multiplicity of pooling equilibria supported by unintuitive beliefs off the equilibrium path. To eliminate such equilibria, the D1 criterion restricts beliefs to be consistent with reasonable inferences, assuming that deviant players' types correspond to the type that would be most tempted by such a deviation. Given this refinement, pooling equilibria with endogenous social bliss points have similar properties to those established in Bernheim (1994)'s analysis. The main result here shows that introducing the endogenous SBP does not induce further multiplicity in the set of equilibria. Bernheim (1994) finds a multiplicity of pooling locations supported by a fixed SBP. Here, each endogenous SBP is supported by a unique pooling location, so the multiplicity of equilibria is entirely driven by the multiplicity of SBP locations.

3.2.1 The Bernheim Results

Define the set of types playing action a, the set's infimum and its supremum as $T(a) = \{t \in T | a^*(t) = a\}$, $t_l(a) = \inf T(a)$, and, $t_h(a) = \sup T(a)$. Since each type has measure zero, Bernheim Theorem 1's monotonicity result implies that $T(a)$ can be written as the closed interval: $T(a) = [t_l(a), t_h(a)]$. Bernheim (1994)'s results characterize pooling equilibria as consisting of a single pool point and a fixed set of types that belong to the pool.

Bernheim Theorem 3 *For fixed α, if $\lambda > \lambda^*$, then for any exogenous social equilibrium that satisfies the D1 criterion, there exists at most one $a_p \in A$ such that $t_l(a_p) < t_h(a_p)$, and it satisfies $\alpha \in T(a_p)$.*

Bernheim Theorem 4 *For fixed α and any given $a_p \in A$, there is at most one central pooling exogenous social equilibrium (a_p, t_l, t_h).*

Outside of the pool, i.e. for agents with types $t \notin T(a_p)$, behavior follows the inference characterizing function discussed in the fully separating equilibrium, so $a^*(t, \alpha) = \phi_s^{-1}(t; \alpha)$.

3.2.2 Implications of Endogenous Social Bliss Point

The Bernheim results identify candidates for endogenous social equilibria, but these candidates do not necessarily satisfy equilibrium condition 3. In fact, only one exogenous social equilibrium satisfies both the D1 criterion and condition 3 for a given SBP.

Theorem 2 *If $\lambda > \lambda^*$, then conditional on the population average strategy, the unique social equilibrium with incomplete separation satisfying the D1 criterion is characterized by a single central pool at $a_p^* = \alpha^* + \varepsilon(\alpha^*, \lambda)$, where:*

$$\varepsilon(\alpha^*, \lambda) = \left(\frac{\int_0^{t_l(a_p)} \left(\alpha^* - \phi_s^{-1}(t; \alpha^*, \lambda)\right) dF(t) + \int_{t_h(a_p)}^{2} \left(\alpha^* - \phi_s^{-1}(t; \alpha^*, \lambda)\right) dF(t)}{P\left(t \in [t_l(a_p), t_h(a_p)]\right)} \right) \tag{9}$$

Theorem 2 identifies the pooling point as a function of the population average action and an additive perturbation. This perturbation is continuous in the pooling point, so a unique equilibrium can be identified where pooling occurrs on the SBP, that is, where $\varepsilon(\alpha^*, \lambda) = 0$.

Theorem 3 *If $\lambda > \lambda^*$, then there exists a unique social equilibrium with incomplete separation satisfying the D1 criterion where the single central pool is located at the social bliss point, i.e., where $\varepsilon(\alpha^*, \lambda) = 0$.*

The most striking result here is that, despite introducing an endogenous mechanism that allows the social bliss point to vary, the set of equilibrium does not increase in any material way. That is, even though the SBP may fall in a wide range, there is only one pooling equilibrium compatible with that SBP. Any other pooling equilibrium supported by the model must correspond to a different location for the SBP.

4 Conclusion

The model explores conformity due to individuals' social interactions despite their varied tastes. Modeling norms as the average action reflects the social desire to be a trend setter whose preferences define norms rather than a mere follower. The conformist outcome is not surprising given Bernheim (1994). The extension is not trivially obvious and the result that this extension introduces no new multiplicity to the set of equilibria helps identify social norms by observed behavior.

Appendix A1: Proofs

Proof of Lemma 1

Given the optimal response function $a^*(t;\alpha,\lambda)$ with a consistent inference function $\phi(b,a;\alpha,\lambda)$ for every $\alpha\in T$, there exists a unique Social Bliss Point, α^*, such that $\alpha^* = E_f[a^*(t;\alpha^*,\lambda,\phi)]$.

Proof. The crux of the argument is showing that $\gamma_\lambda(\alpha) : \alpha \mapsto \int a^*(t;\alpha,\lambda) f(t)\,dt$ is a contraction mapping. Intuitively, the argument is that the degree to which a player's strategy is expected to change in response to a change in the expected mean action of the other players is strictly less than the shift in the mean action. Mathematically, it requires verifying that for all α' and α'' in T:

$$\left|\int a^*(t;\alpha',\lambda) f(t)\,dt - \int a^*(t;\alpha'',\lambda) f(t)\,dt\right| < |\alpha'-\alpha''| \tag{A.1}$$

Lemma A.1 *In any fully separating equilibrium, for all $t\in T$ and all $\alpha'\le\alpha''$, $a^*(t;\alpha',\lambda)\le a^*(t;\alpha'',\lambda)$.*

Proof. The argument follows Bernheim's proof of his Theorem 1, but works on the social utility function h rather than the intrinsic utility function g. Let r be the intrinsic utility associated with choosing $a=a^*(t;\alpha)$ and let r' be the level of intrinsic utility associated with $a'=a^*(t;\alpha')$. Assume $a'>a$ in an equilibrium, which requires: $r+h(a-\alpha)\ge r'+h(a'-\alpha)$, and, $r'+h(a'-\alpha')\ge r+h(a-\alpha')$. Adding these two inequalities gives:

$$h(a'-\alpha')-h(a-\alpha')\ge h(a'-\alpha)-h(a-\alpha) \tag{A.2}$$

Now using the Bernheim trick of applying the Fundamental Theorem of Calculus twice and using the strict concavity of h:

$$\begin{aligned}&[h(a'-\alpha')-h(a-\alpha')]-[h(a'-\alpha)-h(a-\alpha)]\\&=\int_a^{a'} h'(w-\alpha')-h'(w-\alpha)\,dw=\int_a^{a'}\int_\alpha^{\alpha'} h''(w-v)\,dwdv<0\end{aligned} \tag{A.3}$$

This last inequality contradicts A.2 and Lemma A.1 is proved. ■

Lemma A.3 *In any fully separating equilibrium, for any* $t \in T$ *and* $\alpha' \leq \alpha''$*, let* $a' = a^*(t; \alpha', \lambda)$ *and* $a'' = a^*(t; \alpha'', \lambda)$*, then* $a'' - a' \leq \alpha'' - \alpha'$*.*

Proof. This lemma holds simply by evaluating the "income" and "substitution" effects associated with the shift in the social bliss point. First, suppose by contradiction that $a'' - a' > \alpha'' - \alpha'$. From an agent's perspective in (a, b) space, the shift in SBP effectively corresponds to a translation of their utility functions and can be analyzed equivalently to a shift in the budget set (here the inference characterizing function). Here, then, reacting exclusively to the shift in the SBP corresponds to the income effect and can be compensated entirely by exactly translating the inference characterizing function by the same magnitude as the indifference curves. However, such a translation gives rise to substitution effects, and agents will substitute intrinsic utility for social utility. Hence, the only way an agent can overcompensate for a shift in the SBP is if the substitution effect were somehow negative, contradicting the concavity assumptions of both the intrinsic and social utility functions. ■

Lemma A.3 *There is some* $\varepsilon > 0$ *so that:*

$$\left| \int a^*(t; \alpha', \lambda) f(t)\, dt - \int a^*(t; \alpha'', \lambda) f(t)\, dt \right| < |\alpha' - \alpha''| - \varepsilon \tag{A.4}$$

Proof. The proof here revolves around the fixed end points, and proceeds by computing the integrals on the left hand side over the range $[0, \delta)$ and $(2 - \delta, 2]$. Because $\phi_s(0)$ and$\phi_s(2) = 1$, the integral will be strictly less than the value required to make $\varepsilon = 0$:

$$|\alpha' - \alpha''|\, P(t \in [0, \delta) \cup (2 - \delta, 2]) \tag{A.5}$$

The result then follows by Jensen's Inequality.

■

■

Proof of Theorem 1

There exists a unique $\lambda^* > 0$ such that a fully separating equilibrium exists if and only if $\lambda \leq \lambda^*$.

Proof. The result here is fairly direct from the previous theorem and Lemma 1 and the result would be considered more of a corollary but for its organizational role in the paper. As shown in Theorem 2, there exists a $\lambda^* (\alpha) > 0$ for any α such that a separating equilibrium exists if and only if $\lambda \leq \lambda^* (\alpha)$. Connecting this theorem with Lemma 1, that under a complete specification there is exactly one equilibrium social bliss point, α^*, an immediate result is that if $\lambda^* = \lambda^* (\alpha^*)$, a fully separating equilibrium obtains if and only if $\lambda \leq \lambda^*$.

■

Proof of Bernheim Theorems 3 and 4

Bernheim Theorem 3: For fixed α, if $\lambda > \lambda^*$, then for any exogenous social equilibrium that satisfies the D1 criterion, there exists at most one $a_p \in A$ such that $t_l (a_p) < t_h (a_p)$, and it satisfies $\alpha \in T (a_p)$.

Bernheim Theorem 4: For fixed α and any given $a_p \in A$, there is at most one central pooling exogenous social equilibrium (a_p, t_l, t_h).

Proof. Since these theorems are stated for fixed α, the proofs from Bernheim are directly applicable. An extraordinary amount of tedium would be needed to identify and remedy all issues like those addressed in Theorem 2, but nowhere in his analysis is the centrality of the SBP required.

■

Proof of Theorem 2

If $\lambda > \lambda^*$, then conditional on the population average strategy, the unique social equilibrium with incomplete separation satisfying the D1 criterion is characterized by a single central pool at $a_p^* = \alpha^* + \varepsilon (\alpha^*, \lambda)$, where:

$$\varepsilon (\alpha^*, \lambda) = \\ \left(\frac{\int_0^{t_l (a_p)} (\alpha^* - \phi_s^{-1} (t; \alpha^*, \lambda))\, d\pi (t) + \int_{t_h (a_p)}^{2} (\alpha^* - \phi_s^{-1} (t; \alpha^*, \lambda))\, d\pi (t)}{P (t \in [t_l (a_p), t_h (a_p)])} \right)$$

Proof. The result follows immediately by applying equilibrium condition (C) to the intersection of the sets of equilibria established in Bernheim's Theorems (3) & (4). Writing condition (C) in integral form yields:

$$\begin{aligned}\alpha^* =& E_\pi\left[a^*\left(t;\alpha^*,\lambda\right)\right] = \int_T a^*\left(t;\alpha^*,\lambda\right)d\pi\left(t\right)\\ =& \int_0^{t_l} a^*\left(t;\alpha^*,\lambda\right)d\pi\left(t\right) + \int_{t_l}^{t_h} a^*\left(t;\alpha^*,\lambda\right)d\pi\left(t\right) + \int_{t_h}^{2} a^*\left(t;\alpha^*,\lambda\right)d\pi\left(t\right)\\ =& \int_0^{t_l} \phi_s^{-1}\left(t;\alpha^*,\lambda\right)d\pi\left(t\right) + \int_{t_h}^{2} \phi_s^{-1}\left(t;\alpha^*,\lambda\right)d\pi\left(t\right) + \int_{t_l}^{t_h} a_p d\pi\left(t\right)\end{aligned}$$

This computation yields:

$$a_p\int_{t_l}^{t_h} d\pi\left(t\right) = \alpha^* - \int_0^{t_l} \phi_s^{-1}\left(t;\alpha^*,\lambda\right)d\pi\left(t\right) - \int_{t_h}^{2} \phi_s^{-1}\left(t;\alpha^*,\lambda\right)d\pi\left(t\right)$$

, or,

$$\begin{aligned}a_p =& \left(\frac{1}{P\left(t\in\left[t_l,t_h\right]\right)}\right)\left(\alpha^* - \int_0^{t_l} \phi_s^{-1}\left(t;\alpha^*,\lambda\right)d\pi\left(t\right) - \int_{t_h}^{2} \phi_s^{-1}\left(t;\alpha^*,\lambda\right)d\pi\left(t\right)\right)\\ =& \frac{\int_{t_l}^{t_h}\alpha^* d\pi\left(t\right) + \int_0^{t_l}\left(\alpha^* - \phi_s^{-1}\left(t;\alpha^*,\lambda\right)\right)d\pi\left(t\right) + \int_{t_h}^{2}\left(\alpha^* - \phi_s^{-1}\left(t;\alpha^*,\lambda\right)\right)d\pi\left(t\right)}{P\left(t\in\left[t_l,t_h\right]\right)}\\ =& \alpha^* + \frac{\int_0^{t_l}\left(\alpha^* - \phi_s^{-1}\left(t;\alpha^*,\lambda\right)\right)d\pi\left(t\right) + \int_{t_h}^{2}\left(\alpha^* - \phi_s^{-1}\left(t;\alpha^*,\lambda\right)\right)d\pi\left(t\right)}{P\left(t\in\left[t_l,t_h\right]\right)}\end{aligned}$$

■

Proof of Theorem 3

If $\lambda > \lambda^*$, then there exists a unique social equilibrium with incomplete separation satisfying the D1 criterion where the single central pool is located at the social bliss point, i.e., where $a_p = \alpha^*$, or equivalently, $\varepsilon\left(\alpha^*,\lambda\right) = 0$.

Proof. Theorem 3 follows by establishing continuity and monotonicity in α^* of the equation identifying $\varepsilon\left(\alpha^*,\lambda\right)$ and applying the Intermediate Value Theorem to identify a unique point where that equation is zero. Once this continuity is established, all that remains is to show there exists a pooling equilibrium where the pool lies below the SBP and another where the pool lies above the SBP.

$$\varepsilon\left(\alpha^{*},\lambda\right)= \frac{\int_0^{t_l(a_p)}\left(\alpha^{*}-\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)\right)d\pi\left(t\right)+\int_{t_h(a_p)}^{2}\left(\alpha^{*}-\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)\right)d\pi\left(t\right)}{P\left(t\in\left[t_l\left(a_p\right),t_h\left(a_p\right)\right]\right)}$$

$$\begin{aligned}\xi\left(a_p\right)\equiv&\alpha^{*}-a_p\\ =&\int_0^{t_l}\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)d\pi\left(t\right)+\int_{t_h}^{2}\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)d\pi\left(t\right)\\ &+\int_{t_l}^{t_h}a_pd\pi\left(t\right)-\int_0^{2}a_pd\pi\left(t\right)\\ =&\int_0^{t_l}\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)d\pi\left(t\right)+\int_{t_h}^{2}\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)d\pi\left(t\right)\\ &-\int_0^{t_l}a_pd\pi\left(t\right)-\int_{t_h}^{2}a_pd\pi\left(t\right)\\ =&\int_0^{t_l}\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)-a_pd\pi\left(t\right)+\int_{t_h}^{2}\phi_s^{-1}\left(t;\alpha^{*},\lambda\right)-a_pd\pi\left(t\right)\end{aligned}$$

Clearly $\xi\left(a_p\right)<0$ when $a_p<\alpha^{*}$ and $\xi\left(a_p\right)>0$ when $a_p>\alpha^{*}$. Continuity, then, would require that $\xi\left(\alpha^{*}\right)=0$. ■

O.D.E. Characterizing Full Separation in Figure 2

This result uses the spherical example developed in Bernheim, where equation 7 becomes:

$$\phi_s'\left(a;\alpha\right)=-\left(\frac{1}{\lambda}\right)\left[\frac{a-\phi_s\left(a;\alpha\right)}{\alpha-\phi_s\left(a;\alpha\right)}\right] \tag{A.6}$$

To find a critical value that ensures the existence of a fully separating equilibrium, analyze equation (6) as a linear dynamical system in (t,x):

$$\begin{bmatrix} dt/d\tau \\ dx/d\tau \end{bmatrix}=\begin{bmatrix} x-t \\ \lambda\left(\alpha-t\right) \end{bmatrix}=\begin{bmatrix} -1 & 1 \\ -\lambda & 0 \end{bmatrix}\begin{bmatrix} t-\alpha \\ x-\alpha \end{bmatrix}=A\begin{bmatrix} t-\alpha \\ x-\alpha \end{bmatrix} \tag{A.7}$$

In equation A.7, τ is an indexing variable and the existence a fully separating equilibrium is equivalent to the matrix **A** having real eigenvalues, a condition that is entirely independent of the social bliss point (note: this result is unique to the spherical case). As the matrix $\mathbf{A}$ in this setting is identical to Bernheim's, the critical value λ^* for satisfying existence of a fully separating equilibrium is easily seen to be $\lambda^* = {}^1\!/\!{}_4$ for all social bliss points.

References

George A Akerlof. A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics*, 94(4):749–775, 1980.

James Andreoni and B. Douglas Bernheim. Social image and the 50-50 norm: A thoeretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636, 2009.

Abhijit V Banerjee. A simple model of herd behavior. *Quarterly Journal of Economics*, 107 (3):797–817, 1992.

Roland Benabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006.

B Douglas Bernheim. A theory of conformity. *Journal of Political Economy*, 102(5):841–877, 1994.

Sushil Bikhchandani, David Hirschleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5): 992–1026, 1992.

Harold L Cole, George J Mailath, and Andrew Postlewaite. Social norms, savings behavior, and growth. *Journal of Political Economy*, 100(6):1092–1125, 1992.