Conformity Concerns: A Dynamic Perspective

Roi Orzach*

July 18, 2025

Abstract

In many settings, individuals imitate their peers' public decisions for one or both of two reasons: to adapt to a common fundamental state, and to conform to their peers' preferences. In this model, the fundamental state and peers' preferences are unknown, and the players learn these random variables by observing others' decisions. With each additional decision, the public beliefs about these unknowns become more precise. This increased precision endogenously increases the desire to conform and can result in decisions that are uninformative about a player's preferences or perceptions of the fundamental state. When this occurs, social learning about peers' preferences and fundamentals ceases prematurely, resulting in inefficient decisions. In line with findings from social psychology, I identify settings where interventions aimed at correcting misperceptions of the fundamental state have no effect but interventions aimed at correcting misperceptions of peers' preferences lead to more efficient decision-making.

Keywords: Conformity Concerns, Social Learning, Pluralistic Ignorance
JEL Classification Numbers: C72, D83, D90

---

*Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215 (e-mail: orzach@bu.edu).

I thank my advisors Alessandro Bonatti, Glenn Ellison, Robert Gibbons, and Stephen Morris for countless conversations. I additionally thank Chiara Aina, Charles Angelucci, Ian Ball, Abhijit Banerjee, Douglas Bernheim, Inga Deimen, Stefano DellaVigna, Tore Ellingsen, Drew Fudenberg, Ying Gao, Saumitra Jha, Jackson Meija, Navin Kartik, Simone Meraglia, Paul-Henri Moisson, Ellen Muir, Vishan Nigam, Shelli Orzach, Jacopo Perego, Nicola Persico, Kramer Quist, Daniel Rappoport, Rafaella Sadun, Tomasz Sadzik, Lones Smith, Jean Tirole, Jonathan Weinstein, Alexander Wolitzky, as well as participants at the MIT Organizational Economics and Theory groups, North American Summer Meeting of the Econometric Society, Stony Brook Game Theory Conference, and Theoretical Research in Development Economics conference for helpful comments.

1 Introduction

Alcohol abuse remains a major public-health challenge, accounting for a substantial share of worldwide mortality. Initially, prevention programs merely lectured adolescents about the health dangers of drinking; such programs proved ineffective, perhaps because they assumed that consumption decisions depend only on a common fundamental state determining alcohol's health costs (Griffin and Botvin, 2010). Once researchers recognized that choices also hinge on perceptions of peers' preferences, interventions became far more successful: Schroeder and Prentice (1998) show that undergraduates who learned their classmates' true preferences drank roughly 40 percent less than those who received health-risk information alone. This result aligns with evidence that individuals seek social conformity (cf. Bernheim, 1994, for a review) but may systematically overestimate their peers' alcohol preferences (Prentice and Miller, 1993).

Environments where conformity, misperceptions of peers' preferences, and misperceptions of a fundamental state interact are pervasive: political endorsements (Loury, 1994; Geiger and Swim, 2016), female labor force participation (Bursztyn et al., 2020a), and corporate board decisions (Westphal and Bednar, 2005). This paper provides a general framework that explains how these two misperceptions interact with the desire to conform, why both misperceptions may persist, and why some interventions may be more successful than others.

In the model, a community attempts to learn two initially unknown variables: a fundamental state (e.g., the health costs of alcohol consumption) and their peers' average preference type. Learning occurs both privately and socially. Privately, each individual receives a signal about the fundamental state and observes his own preference type, which is informative about his peers' preference types because they are drawn from the same population distribution. Socially, each player observe their peers' decisions, which may be informative about their signals or preference types. In line with the social learning literature (cf. Banerjee, 1992; Bikhchandani et al., 1992), I ask: Given an infinite sequence of decisions, can the public correctly infer these two unknown variables (defined as social learning occurring)? In answering this question, I explicitly model the desire to conform.

Formally, each individual seeks to maximize a utility function that combines their private utility with a conformist utility component. Their private utility is determined by how well their decision adapts to their private preference type and the fundamental state. Their conformist utility, on the other hand, is determined by the conformity concerns (an exogenous parameter common to all individuals) multiplied by the gap between the community's perception of the individual's preference type and the true average preference type of the community. I show that if social learning fails, then in the limit the players face three inefficiencies. First, the players are unable to adapt to their preference types. Further, the players' decisions are based on imprecise beliefs about *both* the fundamental state and the average preference type of their peers.

As I elaborate below, this paper provides a new rationale for why social learning fails: conformity concerns. When conformity concerns are sufficiently small, social learning occurs because the players have heterogeneous preference types (Goeree et al., 2006) and utilities are continuous (Lee, 1993; Ali, 2018; Kartik et al., 2024). However, for high (but, importantly, finite) conformity concerns, social learning fails. In such settings, my model delivers predictions consistent with the empirical literature: social learning fails and effective informational interventions are about peers' preference types. In contrast, the social learning literature would predict that social learning occurs and optimal information interventions are about the fundamental state.

Main Results For most of the analysis, I assume Gaussian uncertainty and linear decision rules; these assumptions are relaxed in Section 6.

In Section 3, I first analyze a benchmark static model where players have common knowledge about both the fundamental state and their peers' average preference type but differ in their own preference types, as described above. Here, conformity concerns place a penalty on the degree to which players adapt to their preference types (cf. Bernheim, 1994). I show that all players choose the same decision independent of their preference types if and only if the conformity concerns exceed a given threshold (Proposition 1).[1] This benchmark shows that when conformity concerns are significant, an individual's decision no longer reflects their private information.

In Section 4, I show that when the conformity concerns are sufficiently large, the key mechanism preventing social learning is that an individual is unable to adapt to his preference type because doing so would incur a large reputational penalty. Further, because adaptation to a player's private signal about the fundamental state would be

[1] I discuss the difference between my benchmark and the environment in Bernheim (1994) in Section 6.1. Reassuringly, in Bernheim (1994) when conformity concerns are sufficiently high, the unique equilibrium is also fully pooling.

attributed to a player's preference type, such adaptations are similarly discouraged. Therefore, when the conformity concerns are large enough, each decision is uninformative of a player's private information. These uninformative decisions imply that if the conformity concerns are sufficiently large, all players pool based on inaccurate perceptions of both the fundamental state and average preference type (Proposition 2). Further, when the average preference type is common knowledge, higher conformity concerns result in greater asymptotic uncertainty about the fundamental state, which in turn implies a lower asymptotic utility (Proposition 3).

Given these two inaccurate perceptions, Section 5 analyzes the different effects of *peer-oriented* interventions, which inform players about their peers' preference types (e.g., how "cool" substance abuse is thought to be by others), versus *individual-oriented* interventions, which inform players about the fundamental state (e.g., the health costs of substance abuse).[2] If individuals have no desire for conformity, they would disregard peer-oriented interventions entirely. Instead, my analysis shows that (i) peer-oriented interventions can break pooling equilibria while individual-oriented interventions can not (Proposition 4), and (ii) individual-oriented interventions may have no effect on which decision the players pool on (Proposition 5).

These findings are consistent with Schroeder and Prentice (1998), Bursztyn et al. (2020a), and Ferreira et al. (2024). Recall, Schroeder and Prentice (1998) showed that peer-oriented interventions lead to 40 percent less alcohol consumption than individual-oriented interventions. Additionally, Bursztyn et al. (2020a) showed that Saudi Arabians misperceived their peers' attitudes towards women working outside the home (WWOH). When these misperceptions were corrected, WWOH increased by 36 percent relative to a control group. Bursztyn et al. (2020a) also argues that information about the economic benefits of WWOH (arguably, an individual-oriented intervention) would have no effect, consistent with my model. Relatedly, Ferreira et al. (2024) shows peer-oriented interventions decrease female genital cutting by 40 percent in Somalia, whereas traditional individual-oriented approaches have largely been ineffective. Finally, misperceptions about a group's average preference type is referred to as "pluralistic ignorance" in social psychology, and I discuss this connection more in Section 4.3.

[2]Throughout, I assume that the party conducting the interventions is benevolent, implying interventions are credible. See Benabou and Tirole (2024) for an analysis where the party conducting the intervention has differing preferences from the community, resulting in a commitment problem.

Related Literature This paper is related to two strands of theoretical literature: social learning and decision-making with reputational concerns. I discuss the empirical literature in Section 4.3.

This paper is closely related to the literature on social learning. Banerjee (1992) and Bikhchandani et al. (1992) show social learning can fail: players inefficiently aggregate information despite observing infinite decisions. However, Lee (1993); Ali (2018); Kartik et al. (2024) show that continuous decisions combined with varying technical conditions are sufficient for social learning to occur.[3] Further, with a similar condition, Goeree et al. (2006) show heterogeneous preference types imply social learning occurs. I allow for both continuous decisions and heterogeneous preference types, and yet find that social learning fails when players possess conformity concerns about their preference types. The social learning literature has documented other obstacles to social learning such as costs of acquiring information (cf. Burguet and Vives, 2000; Chandrasekhar et al., 2018), misspecified priors (cf. Bohren, 2016; Frick et al., 2020), non-bayesian updating (cf. Golub and Jackson, 2010), changing fundamentals (cf. Dasaratha et al., 2023), or differential observability assumptions (cf. Banerjee and Fudenberg, 2004; Arieli and Mueller-Frank, 2019).

The literature on reputational concerns typically analyzes settings where an agent makes an observable decision attempting to both (i) adapt his decision to a signal and (ii) make the observer think the agent is a "good type." Early work, such as Scharfstein and Stein (1990) or Morris (2001), assumed the definition of a "good type" was common knowledge.[4] In contrast, I focus on environments where the decision-maker has multiple observers and the "good type" is different for each observer, such as in: Bernheim (1994), Loury (1994), Manski and Mayshar (2003), Austen-Smith and Fryer Jr (2005), Kuran and Sandholm (2008), Michaeli and Spiro (2015), and Tirole (2023). Further, I explicitly utilize the definition of conformity developed in Bernheim (1994). However, in these papers, the interaction is static and the lone decision-maker maximizes over the distribution of observers. Instead, my dynamic analysis has aggregate uncertainty over the distribution of observers and a key question is how this uncertainty is resolved as the game unfolds.

Finally, within the intersection of these literatures, the impact of conformity con-

[3]For instance, the normal distribution satisfies the "DUB" condition in Kartik et al. (2024).

[4]There exist many related papers which analyze different settings, but continue to assume the "good type" is common knowledge. See Canes-Wrone et al. (2001); Ely and Välimäki (2003); Ottaviani and Sørensen (2006).

cerns on information transmission is considered in Braghieri (2021) and its impact on social learning is considered in Li and Van den Steen (2021) and Fernández-Duque (2022).[5] These models are different from my work because in those papers (i) there is uncertainty about only the players' preference types, not the fundamental state of the world, and (ii) decisions and preference types are discrete. Distinction (i) allows my work to discuss when an individual-oriented or a peer-oriented intervention is preferred. Distinction (ii) allows my work to generate social learning failures where prior models without conformity do not because the literature already predicts social learning may fail with discrete decisions.

2 Model

I consider a community whereby, each period, a player makes a decision attempting to adapt to his private information and private preference type while possessing conformity concerns as described in detail below.

Players: There is an infinite sequence of short-run players, $t \in 1, 2, \ldots$. Each player, t, observes the public history, h_t, (which will be specified after defining the utility), and his private information, and then chooses a decision $a_t \in \mathbf{R}$ in period t. The players possess uncertainty about two random variables, a common fundamental state, θ and the average preference type of the group, μ. This uncertainty takes the following parametrized form,

$$\begin{pmatrix} \theta \\ \mu \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_\theta & \rho\sqrt{\tau_\theta \tau_\mu} \\ \rho\sqrt{\tau_\theta \tau_\mu} & \tau_\mu \end{pmatrix}^{-1} \right), \tag{1}$$

where $\tau_\theta, \tau_\mu > 0$ and $\rho \in (-1, 1)$.[6] Each player's private information is his signal

[5] There is a literature discussing social learning where players value coordination (cf. Angeletos et al., 2007). Bernheim (1994) distinguishes between coordination (wanting to match the actions of one's peers) and conformity (wanting to be perceived as similar to one's peers). Bernheim (1994) shows coordination results in strictly monotone decision rules whereas conformity result in pooling. Additionally, for coordination there should be little difference between private and public decisions, whereas we see differences in Bursztyn et al. (2019, 2020b); Braghieri (2021).

[6] Section 6 discusses generalizations of this framework to other distributions. Further, (i) that the prior means are equal to zero is without loss and (ii) as the beliefs will become correlated in subsequent periods, it is without loss to allow $\rho \neq 0$ ex-ante (see Footnote 11). Finally, to rationalize the empirical findings in Schroeder and Prentice (1998) discussed in the introduction, I allow for uncertainty about θ and μ.

$s_t = \theta + \epsilon_t$ and his preference type $v_t = \mu + \nu_t$. I assume ϵ_t, ν_t are independent within and across players, and that both are Gaussian random variables with means normalized to zero and precisions normalized to one.

Utility: Each player's utility has two components. First, the player wants to adapt his decision, a_t, to a combination of the fundamental state and his preference type. The weight of the fundamental state, $\gamma_t \geq 0$, is publicly observable and discussed further below. Second, while each player observes his own preference type, the player prefers the public's perception of his preference type to be close to the average within the community, which represents the conformity concerns.[7] Define by $\phi(b|h_t, a_t)$ the probability distribution over preference types b that take decision a_t given the public history, h_t. Further, for now, $\phi(\cdot)$ is unconstrained off-path. The total utility for player t is thus,

$$u_t(a_t; v_t, s_t|h_t) := -\mathbb{E}_{\theta,\mu}\Big((a_t - \gamma_t\theta - v_t)^2 + \kappa \int (b-\mu)^2 \phi(b|h_t, a_t)db \Big| v_t, s_t, h_t\Big). \quad (2)$$

The first term in the expectation states that the player wants to choose a decision close to a linear combination of the fundamental state and his preference type. The second term is the conformity term, scaled by $\kappa \geq 0$. The term within the parenthesis is a reduced-form representation of conformity: player t wants the community's perception of his preference type, b, to be close to the true average preference type of the community, μ. Given that the loss function is quadratic, one can show that from the stand-point of which decisions are taken, this is equivalent to player t preferring the inference of his preference type to be close to that of a randomly drawn preference type in the community.

Information: I assume that the public history is $h_t := \{\gamma_1, a_1, \ldots, \gamma_{t-1}, a_{t-1}, \gamma_t\}$. This assumption states that there is full observability of the decisions and when they were made. Given this assumption, one can interpret γ_t as a commonly observed time fixed-effect determining whether the fundamental state or one's preference type is comparatively more important. These time fixed effects are modeled as (i) $\gamma_t \leq 1 \ \forall \ t$, and (ii) there being an $m \in \mathbf{N}$ such that $\gamma_{t+m} = \gamma_t$ and $\gamma_{t+1} \neq \gamma_t$. These assumptions are only necessary in the analysis with uncertainty about θ and μ. Without these assumptions there is only one "moment condition" for the players to separately infer

[7] In the Supplemental Appendix, I show that one would get qualitatively similar results if, throughout the analysis, each player wanted his perceived preference type to be $\mu + c$ with $c \neq 0$.

θ and μ, potentially resulting in incomplete learning. However, in the Supplemental Appendix, I show that such learning outcomes are not locally stable, and this assumption removes their existence.[8]

I analyze Perfect Bayesian Equilibria satisfying the following requirements each period, and I will discuss each requirement below.

1. *Linearity:* Decisions are a linear combination of the public's beliefs about θ, the public's beliefs about μ, a player's private signal, and a player's preference type.

2. *Social Optimality:* In all periods t, the players play the linear equilibrium that maximizes player t's expected utility.

Linearity: The restriction to linear equilibria is common when studying the normal learning model, as it allows for greater tractability. In Section 6, I discuss how these results extend beyond the Linear-Gaussian environment.

Social Optimality: I assume that for each period, t, the equilibrium decision rule in period t maximizes the expected utility out of all linear decision rules in period t. This refinement equivalently selects the equilibrium which maximizes the discounted expected utility of the players from the class of equilibria that are linear in all periods. Without this refinement, then for any sequence $\{x_t\}_{t=1}^{\infty}$, there exists an equilibrium where players pool on x_t in period t. Such equilibria are removed by this requirement, as total utility is higher under equilibria with non-constant decision rules.

These criterion prescribe a unique decision rule in all periods. Thus, I refer to the Perfect Bayesian Equilibrium satisfying these conditions as *the signaling equilibrium*.

3 Common Knowledge Benchmark

This section analyzes the impact of conformity on decision-making and mutes uncertainty about the fundamental state and the preferences of others. To do so, I assume θ and μ are common knowledge, and, without loss of generality, are both equal

[8] These learning outcomes resemble the confounded learning outcomes in Smith and Sørensen (2000). Unlike Smith and Sørensen (2000), these outcomes are not locally stable. Intuitively, this assumption implies that players place different weights on the fundamental state at different times (e.g., after finals with respect to drinking or during election cycles in political speech examples).

to zero.[9] Further, without uncertainty, there is no time dependence, thus it is without loss of generality to consider the decision rule of a player with preference type v. As the decision rule is linear in v, there are two cases: a constant and a strictly increasing decision rule. If the decision rule is constant (respectively, strictly increasing), then this decision rule is defined as "fully-pooling" (respectively, "revealing").

A linear revealing equilibrium is determined by $\mathbb{E}(v|a) := \hat{v}(a) = \alpha a + \beta$. $\hat{v}(a)$ constitutes an equilibrium if and only if given $\hat{v}(a)$, the decision rule that maximizes a players utility, $a(v)$, results in a consistent conjecture of $\hat{v}(a)$. The first-order condition for $a(v)$ given a conjecture $\hat{v}(a) = \alpha a + \beta$ is:

$$a - v + \kappa\alpha(\alpha a + \beta) = 0 \iff v = (1 + \kappa\alpha^2)a + \kappa\alpha\beta. \tag{3}$$

Further, these beliefs result in consistent conjectures if and only if:

$$1 + \kappa\alpha^2 = \alpha \text{ and } \beta = \kappa\alpha\beta. \tag{4}$$

Clearly $\beta = 0$ is a solution to the latter equality, and, further, it is the unique solution for any α that solves the former. While the former equality can be solved directly, Figure 1 depicts the best response, $1 + \kappa\alpha^2$, as a function of the conjectured slope, α, for two different values of κ. As seen in Figure 1, an equilibrium with revelation does not exist for high values of κ. This non-existence follows from two complementary intuitions. (i) As κ increases, the player conforms more: i.e., $a(v)$ has a smaller slope. (ii) In a revealing equilibrium the slope of $a(v)$ cannot be too small. If this occurred, then the slope of the conjectured beliefs, α, would be large. However, when α is too large, the benefit of conformity is so high that each player will choose the same action, implying the equilibrium is not revealing.

Further, Figure 1 provides intuition for why the fully-pooling decision rule is an equilibrium. With full pooling, the slope of the decision rule is zero, implying that the slope of the beliefs (i.e., the inverse of the decision rule) is infinite. Given these beliefs, players choose the pooling decision, resulting in a consistent conjecture.

[9] As θ is common knowledge, players disregard signals s_t. This benchmark is similar to the analysis in Bernheim (1994). Section 6.1 examines this benchmark without assuming linearity or Gaussianity and offers a comprehensive comparison with Bernheim (1994).

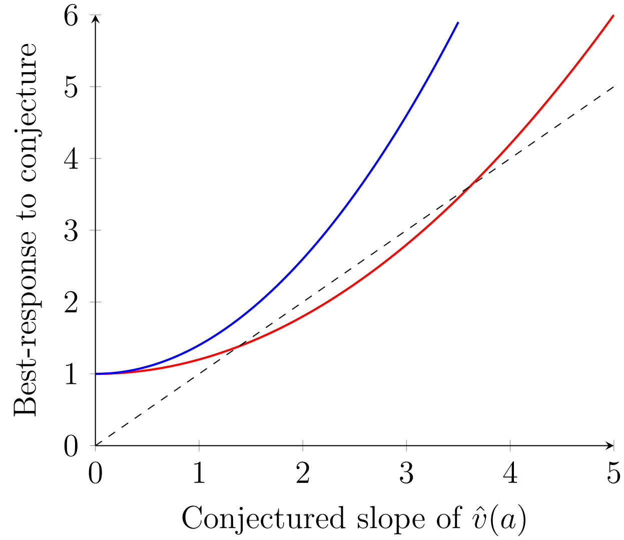

Figure 1: Existence of Linear Equilibria with Revelation

The x-axis represents a conjectured slope of the posterior beliefs of the player's preference type given their decision, $\hat{v}(a)$. The y-axis depicts the beliefs that result from the best response to such a conjecture as stated in Equation (4). In Blue is the best response when the conformity concerns are high ($\kappa = .4$) and in Red the best response when the conformity concerns are low ($\kappa = .2$).

Finally, if the conformity concerns are low enough that a revealing equilibrium exists, then three different linear equilibria exist. In selecting the equilibrium with the highest expected utility, the Law of Iterated Exception implies the conformity loss is fixed across the equilibria, thereby shifting focus to the equilibrium with the minimal adaptation loss. Further, the equilibria with the highest slope of $a(v)$ (equivalently, smallest slope of $\hat{v}(a)$) minimizes the adaptation loss. Therefore, the social optimality refinement selects this equilibrium. These intuitions are formalized below.

Proposition 1 (**Commonly Known Environment**)
There exists a threshold value, $\kappa^{c.k.}$, such that the unique signaling equilibrium is characterized by the following decision rule:

$$a(v) = \begin{cases} \frac{1+\sqrt{1-4\kappa}}{2}v & \text{if } \kappa \leq \kappa^{c.k.} \\ 0 & \text{if } \kappa > \kappa^{c.k.} \end{cases}, \tag{5}$$

where κ denotes the weight on conformity. Given this decision rule,

$$u(v) = \begin{cases} -\frac{1-\sqrt{1-4\kappa}}{2}v^2 & \textit{if } \kappa \leq \kappa^{c.k.} \\ -v^2 - \kappa & \textit{if } \kappa > \kappa^{c.k.} \end{cases}. \tag{6}$$

This proposition formalizes the intuitions from the figure. First, the degree to which a player adapts to his preference type, the slope of $a(v)$, is decreasing with respect to the conformity concerns, κ. Further, when the incentive to conform becomes sufficiently high, i.e., $\kappa > \kappa^{c.k.}$, the signaling equilibrium is fully pooling. Importantly, such fully-pooling decision rules provide no information about a player's private information, and this observation will be key in the main analysis.

4 Analysis

This section begins with an analysis of the complete model where the players attempt to learn both the fundamental state and the average preference type of their peers. The first subsection characterizes the impact of conformity concerns on social learning. The second subsection presents comparative statics on the asymptotic utility and the asymptotic precision of the beliefs about the fundamental state, as defined below, showing that both decrease with κ. In doing so, I assume away uncertainty about the population's average preference type and focus on how the players learn the fundamental state.[10] Incorporating both dimensions of uncertainty allows for predictions that neither uni-dimensional learning model will produce; however, doing so complicates the analysis by requiring a joint update in the posterior beliefs each period. In the final subsection, I relate the predictions of my theoretical model to empirical and qualitative findings.

I now define notation that appears throughout. Define $\theta(t) = \mathbb{E}(\theta|h_t)$ and $\mu(t) = \mathbb{E}(\mu|h_t)$.[11] These random variables have the following joint distribution:

[10] Further, in Subsection 6.2, I conduct an analysis where the fundamental state is common knowledge, but the average preference type, μ, is uncertain and derive similar comparative statics.

[11] Even if $\theta(1)$ and $\mu(1)$ are independent, in subsequent periods, $\theta(t)$ and $\mu(t)$ are dependent as both condition on h_t. At any h_t, $\theta(t), \mu(t)$ are sufficient statistics for the probability distribution determining the players' beliefs. Therefore, I refer to these random variables as "the beliefs."

$$\begin{pmatrix} \theta(t) \\ \mu(t) \end{pmatrix} \sim N \left(\begin{pmatrix} \bar{\theta}(t) \\ \bar{\mu}(t) \end{pmatrix}, \begin{pmatrix} \tau_{\theta,t} & \rho_t \sqrt{\tau_{\theta,t}\tau_{\mu,t}} \\ \rho_t \sqrt{\tau_{\theta,t}\tau_{\mu,t}} & \tau_{\mu,t} \end{pmatrix}^{-1} \right). \tag{7}$$

As is common, it is convenient to work with the precision matrix, defined as the inverse of the variance matrix in Equation (7). Below, I introduce definitions.

Definition 1 (**Social Learning**)
Social learning about fundamentals occurs (respectively, fails) if and only if $\theta(t) \to_p \theta$ (respectively, $\theta(t) \not\to_p \theta$). Social learning about preferences occurs (respectively, fails) if and only if $\mu(t) \to_p \mu$ (respectively, $\mu(t) \not\to_p \mu$).

In the signaling equilibrium, the inference function, $\phi(b|h_t, a_t)$, is Gaussian, the mean of which I denote by $\hat{v}(a_t) := \int b \cdot \phi(b|h_t, a_t)$. Further, the variance of $\phi(b|h_t, a_t)$ and μ are independent of a_t. These observations allow for the following simplification of the utility function up to some constant c_t as stated below.

$$\begin{aligned} u_t(a_t; v_t, s_t|\theta(t), \mu(t), \gamma_t) := &- \mathbb{E}_{\theta,\mu}\Big((a_t - \gamma_t\theta - v_t)^2 \Big| \theta(t), \mu(t), v_t, s_t\Big) \\ &- \kappa\mathbb{E}_{\theta,\mu}\Big(\big(\hat{v}_t(a_t) - \mathbb{E}(\mu)\big)^2 \Big| \theta(t), \mu(t), v_t, s_t\Big) - c_t. \end{aligned} \tag{8}$$

The first term is the adaptation loss. The second term is the squared difference from the expectation of player t's preference type to the average preference type of the population, and the final term corresponds to the variance of $\phi(b|h_t, a_t)$ and μ. Finally, I define the following notions of efficiency.

Definition 2 (**Asymptotic Utility and Adaptation Loss**)
The asymptotic utility is $\lim_{t\to\infty} \mathbb{E}\big(u_t(a_t; v_t, s_t|\theta(t), \mu(t), \gamma_t)\big)$. The asymptotic adaptation loss is $\lim_{t\to\infty} \mathbb{E}\big(-(a_t - \gamma_t\theta - v_t)^2\big)$.

Throughout, I provide results for both notions of asymptotic efficiency. However, the results are qualitatively similar because the difference in these limits is the expected conformity loss, which is pinned down by the Law of Iterated Expectation.

4.1 Determinants of Social Learning

This subsection analyzes the complete model where the players attempt to socially learn θ and μ. To analyze this environment, I consider an arbitrary period, t and recall that $\theta(t), \mu(t)$ are sufficient statistics for the past history.

As the signaling equilibrium is revealing if an equilibrium with revelation exists, I proceed similarly to the benchmark and conjecture a revealing equilibrium with $\hat{v}_t(a_t) = \alpha_t a_t + \beta_t$. Given $\hat{v}_t(\cdot)$, player t's first-order condition for his decision rule is:

$$a_t(1+\kappa\alpha_t^2) = \gamma_t \mathbb{E}\big(\theta \mid \theta(t), \mu(t), s_t, v_t\big) + v_t + \kappa\alpha_t\beta_t + \kappa\alpha_t\mathbb{E}\big(\mu \mid \theta(t), \mu(t), s_t, v_t\big). \quad (9)$$

Further, $\hat{v}(a_t)$ constitutes an equilibrium if the posterior expectation of v_t, given the decision rule in Equation (9), is consistent with the conjecture $\hat{v}(a_t)$. Given the distributional assumptions, one can write this posterior expectation as follows,

$$\mathbb{E}\Big(v_t \mid a_t(1+\kappa\alpha_t^2)\Big) := a_t(1+\kappa\alpha_t^2)\cdot\tilde{s}(\alpha_t, \kappa, \theta(t), \mu(t)) + \iota(\alpha_t, \kappa, \beta_t, \theta(t), \mu(t)), \quad (10)$$

where $\tilde{s}(\cdot)$ will be thought of as determining the sensitivity of the decision rule to v_t and $\iota(\cdot)$ as determining the intercept. Thus a signaling equilibrium is revealing if and only if there exist an α_t and β_t which satisfy,

$$(1+\kappa\alpha_t^2)\tilde{s}(\alpha_t, \kappa, \theta(t), \mu(t)) = \alpha_t \quad (11)$$

$$\iota(\alpha_t, \kappa, \beta_t, \theta(t), \mu(t)) = \beta_t. \quad (12)$$

These equations resemble those in Equation (4) where, in that benchmark, $\tilde{s}(\cdot) = 1$ and $\iota(\cdot) = \kappa\alpha_t\beta_t$. Similar to that benchmark, whenever there exists a solution to Equation (11), there exists a unique solution to Equation (12), thus shifting the focus to Equation (11). Further, Equation (11) is independent of β_t. Finally, the sensitivity, $\tilde{s}(\cdot)$, is independent of the means of the beliefs, $\theta(t), \mu(t)$. This independence arises because the sensitivity captures how variation in the decision corresponds to variation in a player's preference type, which is independent of the mean beliefs in a linear equilibrium. The following lemma formalizes this intuition and provides additional properties of the learning process.

Lemma 1

Fix κ, $\{\gamma_t\}$, and initial beliefs about θ and μ. Social learning about fundamentals occurs if and only if social learning about preferences occurs. Further, whether or not social learning about fundamentals occurs (symmetrically, preferences) is independent of the realizations of a_t.

The intuition behind the first statement in the lemma is that if the players socially learn θ, the players must have observed infinitely many periods of informative decisions. Given the knowledge of what θ is, the players can use these infinitely many periods to infer μ and vise versa. Since the conditions for social learning about preferences and fundamentals are identical, for brevity, I will refer to social learning as when the players socially learn both the preferences and fundamentals.

The second statement formalizes the intuition that whether an equilibrium involves revelation is determined only by the sensitivity, $\tilde{s}(\cdot)$ rather than the intercept, $\iota(\cdot)$. Importantly, the *effective conformity concerns*, defined as the endogenous reputational penalty of adaptation, change over time because the sensitivity is influenced by the precision of the beliefs in a given period. Adaptation imposes a change in one's perceived preference type determined by the sensitivity, and the conformity loss is equal to κ multiplied by this sensitivity. To build intuition, if $\theta(t)$ is sufficiently imprecise, player t puts comparatively more weight on his signal, s_t. The sensitivity is now lower because variation in a_t is ascribed to s_t as opposed to v_t. In contrast, if $\theta(t), \mu(t)$ are sufficiently precise, then player t effectively disregards s_t and v_t when computing the posterior expectations of θ and μ. Therefore, the right-hand side of Equation (9) is approximately equal to v_t, which implies that the sensitivity is approximately equal to 1. One can use this intuition to prove the following lemma, providing a sufficient condition for a failure in asymptotic learning.

Lemma 2 (**Sufficient Condition for Failure of Social Learning**)
If $\kappa > \kappa^{c.k.}$, social learning about fundamentals and preferences fails for any initial beliefs about θ and μ.

This lemma states that when the conformity concerns exceed the threshold for revelation in the common knowledge benchmark, the players are unable to socially learn θ or μ. To gain intuition, note that by Lemma 1 an equilibrium with revelation exists if and only if there exists an α which solves Equation (11). That $\kappa > \kappa^{c.k.} = 1/4$ implies that if the sensitivity were close to one, then there would exist no solution, as the left-hand side of Equation (11) would be strictly greater than the right-hand side for any α. Hence, for the signaling equilibrium to be revealing, the sensitivity cannot converge to one. Further, for social learning to occur, there must exist infinitely many periods of revelation even as the beliefs converge to the truth. However, if the beliefs converge to the truth, the right-hand side of Equation (9) converges to v_t, implying that the sensitivity converges to 1, yielding a contradiction. As a result, when $\kappa > \kappa^{\text{c.k.}}$ the players face three inefficiencies in the limit. First, as the players use a pooling decision rule in the limit, the players are unable to adapt to their preference types. The subsequent two inefficiencies stem from the players utilizing a pooling decision rule based on inaccurate perceptions of *both* θ and μ.

Given Lemma 2, it suffices to analyze $\kappa < \kappa^{\text{c.k.}}$. First note that for any $\kappa < \kappa^{\text{c.k.}}$, there exist initial beliefs such that social learning occurs. As discussed above, for sufficiently precise beliefs, the sensitivity is bounded above by $1 + \epsilon$. By Equation (11), an equilibrium with revelation exists if and only if $1 + \kappa\alpha_t^2$ multiplied by the sensitivity is equal to α_t. Since $\kappa < \kappa^{\text{c.k.}}$, one can increase the left-hand side by ϵ (corresponding to an increase in κ) and there will still exist a solution. Hence, if period one beliefs are sufficiently precise, then period one will be revealing. Further, as period two beliefs are more precise, period two will be revealing, and so on by induction. Finally, the assumed time fixed-effects, $\{\gamma_t\}$, imply that the players can separately identify θ and μ.[12] This intuition implies that if $\kappa < \kappa^{\text{c.k.}}$, there exist sufficiently precise beliefs for which social learning occurs, as formalized below.

Lemma 3 (**Existence of Social Learning**)
If $\kappa < \kappa^{\text{c.k.}}$, then there exists an open set of initial beliefs such that social learning about preferences and fundamentals occurs.

The question now turns to whether $\kappa < \kappa^{\text{c.k.}}$ is a sufficient condition for social learning to occur for all initial beliefs. Intuitively, for social learning to fail, there must exist a period in which no revealing equilibrium exists. That no such equilibrium exists is equivalent to there being no solution to $1 + \kappa\alpha_t^2$ multiplied by the sensitivity in the decision rule to v_t, $\tilde{s}(\cdot)$ equals α_t. This may occur despite $\kappa < \kappa^{\text{c.k.}}$ when $\tilde{s}(\cdot) > 1$. Further, $\tilde{s}(\cdot) > 1$ may occur when, given a sequence of high decisions, each player is unsure if the decisions were high due to a high θ or a high μ. If v_t is low, he updates that μ is low and thus θ is high, as $\mu(t)$ and $\theta(t)$ are negatively correlated. Therefore, as players negatively weight v_t in their posterior expectation of θ, an ϵ change in a_t corresponds to a greater than ϵ change in v_t, implying $\tilde{s}(\cdot) > 1$. The formal proof showing that that social learning may fail despite $\kappa < \kappa^{\text{c.k.}}$ involves additional complications, as the sensitivity is endogenous to the conjectured beliefs. The following proposition formalizes this logic and unifies the previous lemmas.

[12] Absent this assumption the asymptotic learning outcomes may be unstable (see Footnote 8).

Proposition 2 (**Characterization of Long-Run Learning**)
There exists a threshold $\underline{\kappa} \in (0, \kappa^{c.k.})$ *such that:*

1. *If* $\kappa \leq \underline{\kappa}$, *then for any initial beliefs, social learning about preferences and fundamentals occurs.*

2. *If* $\kappa \in (\underline{\kappa}, \kappa^{c.k.})$, *then there is an open set of initial beliefs such that social learning about preferences and fundamentals occurs. Further, there exists an open set of parameter values for which social learning about preferences and fundamentals fails despite* $\kappa \in (\underline{\kappa}, \kappa^{c.k.})$.

3. *If* $\kappa > \kappa^{c.k.}$, *then for any initial beliefs, social learning about preferences and fundamentals fails.*

The first statement can be seen by analyzing the condition for the existence of a signaling equilibrium with revelation. If κ is sufficiently small, then for any prior beliefs one can show that the conformity concerns are sufficiently small such that there exists an equilibrium with revelation in every period. Further, the third result is a direct consequence of Lemma 2. The second result states that when the conformity concerns take an intermediate value, social learning may occur or fail. That social learning may occur is a consequence of Lemma 3. That social learning may fail is due to the effective conformity concerns being larger than the conformity concerns when there is a sufficiently strong negative correlation between the beliefs about θ and μ.

This section shows that the condition for social learning-about either fundamentals or preferences-to occur depends on an intuitive fundamental: the magnitude of conformity concerns.[13] I discuss this finding in the context of the empirical literature in Section 4.3.

4.2 Comparative Statics

This subsection shows the result that social learning about fundamentals occurs if and only if conformity concerns are sufficiently small continues to hold absent uncertainty about the average preference type, μ, and presents additional comparative

[13] At this level of generality, the learning outcomes are not monotone in κ. For instance, a marginally higher value of κ alters whether the decision rule places marginally more weight on s_t or v_t, which results in different beliefs in the subsequent period. In either analysis with only one dimension of uncertainty, the learning outcomes are strictly monotone with respect to κ.

statics (Section 6.2 contains the parallel analysis). Therefore, I assume μ is common knowledge and without loss of generality equal to zero in the following proposition. Further, as the learning is univariate, one can normalize $\gamma_t = 1$ without loss of insights (see Footnote 8).

Proposition 3 (**Social Learning when Average Preference Type is Known**) *Fix any prior beliefs about θ and let $\mu = 0$ be common knowledge. Social learning about fundamentals occurs if and only if $\kappa \leq \kappa^{c.k.}$. Further, when $\kappa > \kappa^{c.k.}$ the long-run precision of the beliefs $\tau_\theta(\kappa) := \lim \tau_{\theta,t}(k) < \infty$ is decreasing in κ. Finally, the asymptotic adaptation loss and asymptotic utility of the players is decreasing in κ.*

This proposition shows that conformity concerns impacts not only the binary outcome of social learning, but also the degree of asymptotic learning. The intuition is that greater uncertainty about θ implies that deviations in a_t are increasingly ascribed to s_t as opposed to v_t. As a result, the greater the uncertainty, the lesser the effective conformity concerns, making it easier to sustain an equilibrium with revelation. Therefore, if the equilibrium is revealing in the common knowledge benchmark, then the equilibrium is revealing in all periods with uncertainty about θ. This implies that if $\kappa \leq \kappa^{c.k.}$, then the players learn θ for any initial belief. However, if $\kappa > \kappa^{c.k.}$, then in the common knowledge environment the signaling equilibrium is pooling, implying that for sufficiently precise beliefs the players pool, thus precluding social learning.

Further, when the conformity concerns are higher, the players switch to the pooling equilibrium earlier. Given this earlier switch, an increase in κ when $\kappa > \kappa^{\text{c.k.}}$ results in pooling on less accurate perceptions of θ, ultimately resulting in both a worse asymptotic adaptation loss and a worse asymptotic utility. In contrast, for $\kappa \leq \kappa^{c.k.}$, the asymptotic utility of the players converges to that in the common knowledge benchmark. The following figure showcases this intuition, by plotting the asymptotic adaptation loss as a function of κ in both the benchmark (blue) and the dynamic model (red). Here, the discontinuity at $\kappa = \kappa^{\text{c.k.}}$ occurs because the players can adapt to their private type in the limit if and only if $\kappa \leq \kappa^{\text{c.k.}}$.

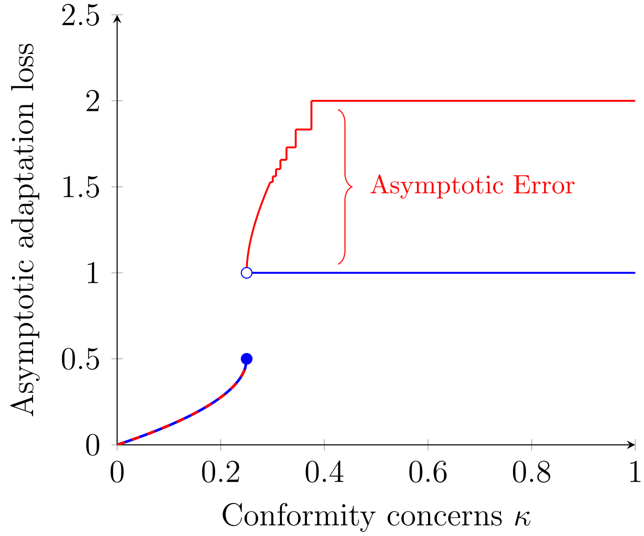

Figure 2: Asymptotic Adaptation Loss

In Blue is the asymptotic adaptation loss in the benchmark. In Red is the asymptotic adaptation loss when μ, the average preference type, is common knowledge, but have prior $\theta \sim N(0, 1)$ about the fundamental state. Finally, the asymptotic errors equals $1/\tau_\theta(\kappa)$, as defined in Proposition 3.

This simplified model showcases how conformity concerns impede efficient learning. Despite the players having access to continuous decisions and sufficiently informative signals, the players fail to learn the fundamental state when the conformity concerns are sufficiently high. Further, the conformity concerns effect the intensive margin of the precision of the beliefs: the greater the conformity concerns, the more imprecise beliefs the players ultimately harbor.

4.3 Applied Relevance

In this subsection, I first argue that my model's learning predictions are more consistent with the empirical literature than the existing theoretical literature. Next, I discuss connections with the literature on pluralistic ignorance.

Determinants of Efficient Decisions: In my analysis, the magnitude of conformity concerns is the main predictor for whether decisions will be asymptotically efficient. In contrast, the social learning literature predicts that continuous decisions or unboundedly informative signals are sufficient for asymptotic efficiency. I now review the empirical literature in support of my predictions.

Continuous Decisions: Lee (1993) argues that continuous decisions imply efficient decisions asymptotically and uses financial markets as an exemplar where investment decisions are continuous, and, indeed, typically firms learn whether a given asset is valuable. In contrast, my model predicts that the conformity concerns must be low for efficient decisions asymptotically. For financial markets, one might think the conformity concerns are low relative to the financial stakes, implying my model also predicts efficient asymptotic decisions. However, if decisions are continuous, but the conformity concerns are high, such as alcohol consumption (cf. Prentice and Miller, 1993), drug use (cf. West and O'Neal, 2004), and many others, then in line with my model, we see more inefficiencies and inaccurate beliefs amongst the community.

Unboundedly Informative Signals: The social learning literature defines a signal as unboundedly informative if with positive probability a player's private signal renders the player arbitrarily certain of the optimal decision. A common result is that if these signals occur, social learning occurs: if the players are herding on a wrong decision, eventually such a signal occurs, breaking the herd. This result is in contrast to the famous conformity experiment in Asch (1953). Participants were grouped and shown a series of lines, then asked to identify the one matching a reference line. Unbeknownst to the participants, the experimenters planted an actor into the group to deliberately provide incorrect answers. Without the actors, success exceeded ninety-nine percent, but with the actors, over seventy-five percent of participants conformed. The social learning literature predicts that individuals should not conform because each individual can identify the correct answer. In contrast, my model predicts that if conformity concerns are large, the individuals will conform. Finally, Franzen and Mader (2023) replicated Asch (1953) and found that monetary incentives decrease the probability of conformity by 13 percentage points. Consistent with my model, these incentives increase the importance of adaptation, resulting in less conformity.

Pluralistic Ignorance: My model predicts that "pluralistic ignorance" can arise in equilibrium. Miller and Prentice (1994) defines pluralistic ignorance as, "a situation in which group members systematically misestimate their peers' attitudes." In a review article, Bursztyn and Yang (2022) document that such misperceptions lead to inefficient social norms and occur in a wide range of environments: political movements, macroeconomic expectations, vaccination status, and many others. Additionally, Miller (2023) argues that pluralistic ignorance stems from two forces, "(1) social life depends on individuals having knowledge of their peers' habitual feelings and practices, and (2) individuals must infer this knowledge from limited and thus potentially misleading information." In my model, these two forces correspond to (1) players having conformity concerns and (2) finitely many periods with informative decisions.

The extent of pluralistic ignorance corresponds to the magnitude by which individuals misestimate the preference types of their peers for a given realization of their peers' true preference types.[14] In the model, if $\mu(t) \to \mu$, then (tautologically) every player correctly predicts μ. In contrast, if $\mu(t) \not\to \mu$, then the uncertainty about μ implies that each player's estimate of μ combines both $\mu(t)$ and his preference type, v_t, which is predictive about μ. Further, when the variance of $\mu(t)$ is higher, then each player's expectation of μ is closer to the prior mean of μ as opposed to the realization of μ. Thus, one can view the variance of $\mu(t)$ (the inverse of $\tau_{\mu,t}$) as a measure of pluralistic ignorance.

While there exist numerous behavioral explanations for pluralistic ignorance, the model presented in this paper provides an additional explanation: the desire for conformity necessitates self-censorship in public discourse (cf. Loury, 1994; Braghieri, 2021), resulting in insufficient information for others to gauge the views of the public.[15]

5 Policy Interventions

In this section, I extend the model to analyze informational interventions and show that when conformity concerns are high, interventions addressing misperceptions of the average preference type outperform interventions addressing misperceptions of the fundamental state.

[14]In the model, $\mu(t)$ is an unbiased estimator for μ. However, the object of interest is the gap between $\mu(t)$ and μ for a given realization of the preferences, μ. One may object that, in practice, $\mu(t)$ is greater than μ in every school in the context of alcohol (rather than being unbiased). However, all these students observe similar sets of celebrities on social media or television drinking, implying that the beliefs across schools should not be viewed as independent samples.

[15]Braghieri (2021) documents that participants skew their answers to politically sensitive questions in the direction of public support, thereby decreasing information transmission. In contrast, I allow for uncertainty about the direction of public support, resulting in a tighter connection with the literature on pluralistic ignorance.

5.1 Modeling Interventions

I consider four types of interventions composed of the intersection of whether the information shared with individuals is made common knowledge and whether the information is about the fundamental state or the average preference type.

Before analyzing "common-knowledge" interventions, I analyze "private interventions." One can think of a private intervention as giving player t access to additional information; however, such information is private and is not accounted for by the community when inferring player t's preference type given his decision. Without formally stating the result, one can see that such an intervention has no ability to break a pooling equilibrium nor influence which decision the players pool on.[16] To see why, suppose player t is told the value of the fundamental state, θ. Given that each player wants to match his decision to the fundamental state, player t has an identical incentive to adapt to θ as a hypothetical player who received a s_t such that $\mathbb{E}(\theta|h_t, s_t)$ equals θ. Further, since such an s_t exists in the support of possible realizations, and that hypothetical player cannot adapt to such information, neither can player t. Thus, the equilibrium in period t remains identical. Finally, as such information was private information to player t, and no change in behavior occurs in period t, then no change in behavior follows for any subsequent periods.

Given the stark irrelevance result for private interventions, I now focus on common knowledge interventions. In the standard framework absent interventions, the public history at time t is $\mathbf{h}_t = \{\gamma_1, a_1, \ldots, \gamma_{t-1}, a_{t-1}, \gamma_t\}$, namely the sequence of past decisions and the environments in which such decisions were chosen. I consider an intervention where information is released before period t, but after a_{t-1}. Such an intervention leaves the prior histories unchanged (and further the prior sequence of events remains unchanged as each player is short-lived). This information could be about either θ or μ, which will be referred to as individual-oriented and peer-oriented interventions, respectively.

5.2 The Effects of Interventions

I begin with a definition of fragility. I call a pooling decision rule "fragile" to an individual-oriented intervention with n pieces of information if a hypothetical

[16] In support of this theory, Tevyaw et al. (2007) shows that the reduction in alcohol use for common-knowledge interventions was 3 times larger than private interventions.

public disclosure of n independent signals that are distributed identically to s_t cause the equilibrium in period t to be non-pooling when it would otherwise be pooling. Similarly, it is fragile to a peer-oriented intervention with n pieces of information if n independent signals distributed identically to v_t cause the equilibrium to be non-pooling when it would otherwise be pooling. This definition extends the definition of fragility in Bikhchandani et al. (2021) to signals about μ.

Proposition 4 (**Fragility**)
The following are true:

1. *If $\kappa < \kappa^{c.k.}$, for any pooling equilibrium there exists an N such that the pooling equilibrium is fragile to both an individual-oriented intervention and to a peer-oriented intervention with $n \geq N$ pieces of information. Further, after either intervention, social learning about fundamentals and preferences occurs.*

2. *If $\kappa > \kappa^{c.k.}$, for any $n \in \mathbf{N} \cup \infty$, an individual-oriented intervention (respectively, peer-oriented intervention) with n pieces of information will never result in social learning about μ (respectively, θ).*

3. *If $\kappa > \kappa^{c.k.}$ and the correlation between the beliefs about θ and μ is equal to zero, the equilibrium is never fragile to an individual-oriented intervention with n pieces of information but may be fragile to a peer-oriented intervention with n pieces of information.*

The intuition behind the first result is that if $\kappa < \kappa^{c.k.}$, then when there is common knowledge about θ and μ the signaling equilibrium involves revelation. Further, that the equilibrium is pooling despite a $\kappa < \kappa^{c.k.}$ is necessitated by a strong negative correlation between the beliefs about θ and μ (Proposition 2). A sufficiently large amount of information about either θ or μ will weaken the negative correlation in these beliefs, allowing for a revealing equilibrium.

The second result says that if κ is high, then giving information about only one dimension of uncertainty will be unsuccessful in spurring social learning on the other. This result follows directly from Proposition 2, which states that for any beliefs about θ and μ, if $\kappa > \kappa^{\text{c.k.}}$, the players' beliefs about θ and μ cannot converge to the truth. Therefore, even if a social planner designs a perfect individual-oriented intervention, the players will necessarily continue to have misperceptions about μ.

Finally, the intuition for the final result comes from the different effects of these interventions on the effective conformity concerns. If there is no correlation between the beliefs about θ and μ, then the player's first-order condition in Equation (9) is

$$a_t(1+\kappa\alpha^2) = \gamma_t\mathbb{E}(\theta|\theta(t), s_t) + v_t + \kappa\alpha\mathbb{E}(\mu|\mu(t), v_t). \tag{13}$$

An individual-oriented intervention always decreases the weight players place on s_t, thus making the decision rule more sensitive to v_t. This increased sensitivity implies that giving information about θ magnifies the effective conformity concerns and thus cannot break a pooling equilibrium.

In contrast, a peer-oriented intervention has two effects. First, each player wants to be perceived as μ. Consequently, when $\tau_{\mu,t}$ is low, players with different preference types have different perceptions of μ. Therefore, each player has an additional reason to adapt a_t to v_t: a high v_t indicates a high μ, implying player t wants to be perceived as a high type, resulting in an incentive to choose a higher a_t. This logic implies that when $\tau_{\mu,t}$ is low, the players have an added incentive to adapt, resulting in revelation.

The countervailing effect is that when $\tau_{\mu,t}$ is low, the uncertainty over a given player's preference type is also high. As is standard in signaling games, when the uncertainty over a given player's preference type is higher, the same player has a greater incentive to signal, and thus a lower incentive to adapt.

Note that the relative value of $\tau_{\theta,t}$ has no affect on the first force but does effect the latter. To see why $\tau_{\theta,t}$ impacts the latter force, note that when $\tau_{\theta,t}$ is high, each decision is mostly determined by a player's preference type, v_t, and not their signal, s_t. As the decision is primarily a function of v_t, a sufficiently precise signal of v_t is generated. This precise signal implies the community's inference about v_t is less sensitive to changes in the prior, such as an increase in $\tau_{\mu,t}$. As a result, when $\tau_{\theta,t}$ is high, increasing $\tau_{\mu,t}$ has a comparatively small effect on the community's inference about v_t and a comparatively large effect on player t's inference about μ. This intuition is seen in Figure 3 below: when $\tau_{\mu,t} > \tau_{\theta,t}$ (respectively, $\tau_{\mu,t} < \tau_{\theta,t}$) an increase in $\tau_{\mu,t}$ causes a change to a revealing equilibrium (respectively, pooling equilibrium).

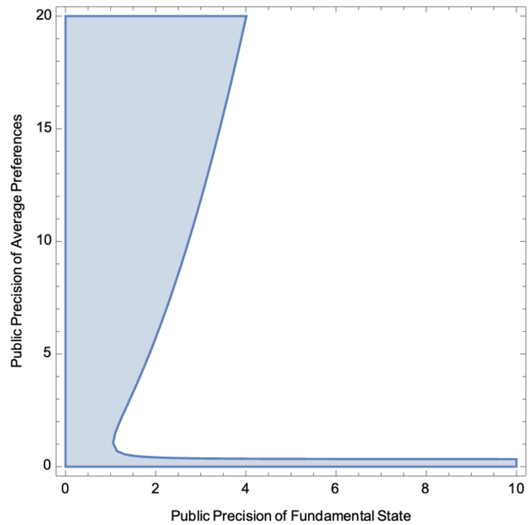

Figure 3: When the Signaling Equilibrium Involves Revelation

In the figure, the x-axis corresponds to $\tau_{\theta,t}$, the public precision of the fundamental state, and the y-axis corresponds to $\tau_{\mu,t}$, the public precision of the average preference type. The shaded region (respectively, non-shaded) corresponds to when the signaling equilibrium is revealing (respectively, pooling). In the figure $\kappa = \gamma_t = 1$, which correspond to the weight of conformity and the relative value players place on adapting to the fundamental state.

However, even if neither a peer-oriented intervention nor an individual-oriented intervention break a pooling equilibrium, these interventions will influence which decision the players pool on. To gain intuition into the forces behind the change, I consider the following situation: suppose $\mu(1), \theta(1)$ are uncorrelated, the equilibrium in period 1 was revealing, and thereafter the equilibrium is pooling. Recall a_1 denotes the decision chosen in period 1. This decision influences what the pooling decision will be in period 2, $a^*(a_1)$, where for simplicity I assume $\gamma_2 = 1$ to derive:

$$a^*(a_1) = \mathbb{E}(\theta|a_1) + \mathbb{E}(\mu|a_1). \tag{14}$$

Further, recall that a_1 is a linear combination of both player one's private signal about θ, s_1, and player one's preference type, v_1, yielding:

$$a_1 = \lambda_\theta s_1 + \lambda_v v_1, \tag{15}$$

where λ denotes such weights. I now consider the following intervention where the public history is adapted to be either $\mathbf{h}_2(\theta) = \{a_1, \gamma_1, \theta\}$ or $\mathbf{h}_2(\mu) = \{a_1, \gamma_1, \mu\}$ and analyze the change in the pooling decision rule that follows. Suppose that the players utilize the decision-rule in Equation (15), resulting in a pooling decision rule denoted by $a^*(a_1)$ as in Equation (14) for all subsequent periods.

Proposition 5 (**Interventions**)
Denote by $\Delta(\theta)$ (respectively, $\Delta(\mu)$) as the difference between the new decision the players pool on compared to $a^*(a_1)$. Then,

$$\Delta(\mu) = \mu\left(1 - \frac{\lambda_\theta \lambda_v}{\frac{1+2\tau_\theta}{1+\tau_\theta}\lambda_\theta^2 + \frac{\tau_\theta}{1+\tau_\mu}\lambda_v^2}\right) + \alpha_\mu a_1 \tag{16}$$

$$\Delta(\theta) = \theta\left(1 - \frac{\lambda_\theta \lambda_v}{\frac{1+2\tau_\mu}{1+\tau_\mu}\lambda_v^2 + \frac{\tau_\mu}{1+\tau_\theta}\lambda_\theta^2}\right) + \alpha_\theta a_1, \tag{17}$$

for some constants α_θ, α_v.

To understand the expressions above, let us now consider the effect of an individual-oriented intervention revealing θ, where a symmetric analysis occurs for μ. Tautologically, $\mathbb{E}(\theta|\mathbf{h}_2(\theta)) = \theta$. Further, the players re-evaluate the perception of μ as a function of both θ (the second term in the parentheses of Equation (17)) and a_1. The object of interest is how much the players decision changes with respect to μ. One can see that given information that θ is positive (respectively, negative) the players update that μ is negative (respectively, positive). Further, this update could be larger in magnitude than the update about the value of θ. Specifically, these cases occur when τ_μ is small (i.e., the players are uncertain about their peers' true preferences). Such a counter-update provides one rationale why the individual-oriented interventions have small (if not negative) effects on behavior.

5.3 Designing Effective Interventions

While both interventions have their merits in different circumstances, the model predicts differential effectiveness. In the model, when conformity concerns are large the players enter into a pooling equilibrium based on inaccurate perceptions of their peers' preference types and the fundamental state. Proposition 4 suggests that peer-oriented interventions may be preferred due to their ability to break a pooling equilibrium. Further, Proposition 5 suggests that even a perfect individual-oriented intervention alone may fail to shift the pooling decision in the direction of efficiency.

These predictions are broadly consistent with the results in Schroeder and Prentice (1998), Bursztyn et al. (2020a), and Ferreira et al. (2024) for two reasons. First, interventions addressing misperceptions of peers' preference types are preferred. Second, in these settings conformity concerns are arguably high. If instead conformity concerns were low (or equal to zero), then similar to the theoretical literature, the optimal interventions are individual-oriented. In such cases, knowledge of the average preference type is less decision-relevant and individual-oriented interventions allow the players to reach an efficient decision faster.

6 Extensions

In this section I consider two extensions that serve as robustness checks for the assumptions in the main analysis. First, I show that non-linear decision rules fail to exist in the benchmark environment in Section 3 if off-path beliefs satisfy D1 from Cho and Kreps (1987). Next, I discuss how the results of the model generalize to alternative distributional assumptions with non-linear decision rules.

6.1 Non-Linear Equilibria

This subsection analyzes Perfect Bayesian Equilibrium when θ and μ are common knowledge, as in Section 3. Similarly to Section 3, it suffices to consider a static version of the game and drop any time-dependence reducing Equation (2) to:

$$u(v, a) = -(a - v)^2 - \kappa \int b^2 \phi(b|a) db. \tag{18}$$

Furthermore, the linear equilibria derived in Section 3 continue to exist for any full-support distribution of v that is atomless, which will be assumed throughout the remainder of this subsection. Throughout, I impose D1, which implies that off-path $\phi(\cdot)$ is concentrated on the types who have the largest incentive to deviate to such a decision. The following lemma defines a "central pooling equilibrium" and shows that any equilibrium which satisfies D1 is a central pooling equilibrium.

Lemma 4 (**Class of Equilibria**)
Any equilibrium satisfying D1 is a central pooling equilibrium. A central pooling equilibrium satisfies $a(v) = c^* \quad \forall v \in [\underline{v}, \bar{v}]$ where $\underline{v} \leq 0 \leq \bar{v}$. Further, if $v \notin [\underline{v}, \bar{v}]$, then

$a(v)$ is continuously differentiable with a derivative that satisfies:

$$a'(v) = \frac{\kappa v}{v - a(v)} > 0. \tag{19}$$

For the proof, I refer the reader to Theorem 3 in Bernheim (1994). The intuition behind his result is that D1 implies $a(v)$ is monotone. Given monotonicity, one can show that D1 further implies that a jump discontinuity cannot arise outside of the central pool. Finally, outside the central pool one can show $a(v)$ is strictly monotone, implying a well-defined inverse of $a(v)$. This inverse can be substituted in for $\phi(b|a)$ to generate the differential equation in Equation (19).

Inspecting Equation (4), one can see that there are two linear equilibria with revelation. Therefore, one can construct a non-linear equilibrium as follows: if $v \geq 0$, utilize the decision rule of one such equilibrium and if $v \leq 0$, utilize the decision rule of the other equilibrium. This equilibrium exists whenever the linear equilibria exist because the problem is symmetric about $v = 0$; however, this equilibrium provides a lower utility than the signaling equilibrium for each v. The following proposition shows these are the only non-linear equilibria satisfying D1. Therefore, up to symmetry about $v = 0$, equilibria satisfying D1 are linear.[17]

Proposition 6 (**Linear Equilibria**)
An equilibrium with revelation satisfying D1 exists if and only if $\kappa \leq \kappa^{c.k.}$. Any such equilibrium has an empty central pool and, on either side of this central pool, the decision rule equals that of a linear equilibrium. However, for any κ there exists a fully-pooling, and hence linear, decision rule.

This result gives support for the restriction to linear equilibria in the main analysis.

6.2 General Distributions

Throughout the analysis I focused on the Gaussian distribution and linear equilibria. When the uncertainty is non-Gaussian, beliefs will be non-linear in a player's private information, resulting in non-linear equilibria. In this subsection, I discuss to

[17]The primary difference to Bernheim (1994) is that I assume that the support of v equals the real line as opposed to a bounded interval. Therefore, in Bernheim (1994) at the boundary of the support $a(v) = v$, whereas in my setting $a(v) \neq v$ outside the central pool. This difference generates the non-linearity and equilibria with partial revelation in Bernheim (1994). Reassuredly, in Bernheim (1994) when conformity concerns are sufficiently high, the unique equilibrium is also fully pooling.

what extent my results generalize to different distributions. As the complete environment with general distributions and both dimensions of uncertainty is intractable, I conduct two separate analyses, each focusing on a different dimension of uncertainty. In this subsection, I assume θ is common knowledge and the players are learning μ, and the Supplemental Appendix contains the other analysis and a discussion of the resulting challenges if both μ and θ are unknown.

Denote by $g_t(v_t|h_t) := \mathbb{E}(\mu|v_t, h_t)$, which, with an abuse of notation, will be denoted as $g_t(\cdot)$. I make two assumptions (i) the distribution of v_t is full support and atomless and (ii) $g_t'(v_t) \in (0, 1)$ and is continuous. Given these assumptions, an immediate generalization of Lemma 4 is that the only equilibria satisfying D1 are central pooling with a derivative outside the central pool satisfying:

$$a_t'(v_t) = \frac{\kappa\big(v_t - g_t(v_t)\big)}{v_t - a_t(v_t)}. \tag{20}$$

Equipped with this differential equation, one can show the following result.

Proposition 7 (**General Distributions with θ known**)
Denote by $\iota_t := \inf_x g_t'(x) \le \sup_x g_t'(x) := \xi_t$, where $g_t(x) = \mathbb{E}(\mu|x, h_t)$. There exists a fully revealing equilibrium that satisfies D1 if the conformity concerns, κ, satisfy

$$\kappa \le \frac{\kappa^{c.k.}}{1 - \iota_t}. \tag{21}$$

Further, no equilibrium with revelation satisfies D1 if

$$\kappa > \frac{\kappa^{c.k.}}{1 - \xi_t}. \tag{22}$$

This proposition gives a separate necessary and sufficient condition for the existence of an equilibrium with revelation. In the linear Gaussian analysis, $g_t(v_t)$ is linear for all t implying that $\xi_t = \iota_t$ and that these bounds are tight. Further, in such an analysis ξ_t is deterministic and decreases if and only if the previous period involved revelation. This determinism implies a monotone relation between κ and the asymptotic precision in the beliefs about μ.

Beyond the linear Gaussian framework, ξ_t, ι_t need not be equal nor monotone in t, precluding such comparative statics. However, one can view ξ_t, ι_t as some measures of asymptotic uncertainty, and Proposition 7 bounds these asymptotic measures.

Further, a corollary to Proposition 7 is that $\kappa \leq \kappa^{\text{c.k.}}$ implies that a fully-revealing equilibrium exists every period. Therefore, if such an equilibrium is selected, then $\mu(t) \to_p \mu$ by the Law of Large Numbers. Further, if $\kappa > \kappa^{\text{c.k.}}$, then $\xi_t \not\to_p 0$, implying $\mu(t) \not\to_p \mu$. These results imply that even beyond the linear-gaussian environment, the magnitude of conformity concerns continues to predict learning outcomes.

7 Conclusion

This paper studies how conformity concerns impact social learning and what interventions are effective when social learning fails. To do so, I enrich a standard model of social learning by adding: (i) a player's desire to adapt to not only a fundamental state but also his private preference type, (ii) an assumption that players have conformity concerns over how the community perceives their private preference type, and (iii) an assumption that there is aggregate uncertainty about the distribution of private preference types in the population. I show that as the players' beliefs about the fundamental state become more precise, the equilibrium penalty experienced by a player who adapts to his private information or his private preference type increases, creating endogenous self-censorship. Further, I show that if the initial conformity concerns are sufficiently high, the endogenous self-censorship not only dampens but eliminates the player's adaptation, resulting in a switch from a revealing to a pooling equilibrium in finite time. Such a switch to pooling implies that forever after the players hold imprecise beliefs about both the fundamental state and the preference types of their peers; the latter is a common finding in social psychology, defined as pluralistic ignorance. Not only are the players pooling (and thus unable to adapt to their private preference types), they pool on an inefficient decision based on imprecise beliefs. Finally, information about the fundamental state has a lower ability to break a pooling equilibrium than information about peers' preferences. This result provides a framework to formalize intuitions extensively discussed empirically in social psychology and economics.

This paper introduced a theoretical methodology that can be used to analyze pluralistic ignorance and how decisions change upon dispelling pluralistic ignorance. I hope this framework can be used to analyze related topics in the social sciences. For instance, related to pluralistic ignorance, there is a large literature on "false polarization" whereby individuals of two distinct subgroups will incorrectly perceive the preferences of the two groups as further apart than reality. Further, there exist numerous empirical studies on “risky and cautious shifts,” whereby upon learning whether the members in their group have risky (respectively, cautious) opinions, the opinions of the group will shift to be more polarized than the opinions of the group members themselves (cf. Sunstein, 2009).

References

Ali, S Nageeb, “On the role of responsiveness in rational herds,” *Economics Letters*, 2018, *163*, 79–82.

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan, “Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks,” *Econometrica*, 2007, *75* (3), 711–756.

Arieli, Itai and Manuel Mueller-Frank, “Multidimensional social learning,” *The Review of Economic Studies*, 2019, *86* (3), 913–940.

Asch, Solomon E, “Effects of group pressure upon the modification and distortion of judgements,” *Group dynamics*, 1953.

Austen-Smith, David and Roland G Fryer Jr, “An economic analysis of ‘acting white’,” *The Quarterly Journal of Economics*, 2005, *120* (2), 551–583.

Banerjee, Abhijit and Drew Fudenberg, “Word-of-mouth learning,” *Games and economic behavior*, 2004, *46* (1), 1–22.

Banerjee, Abhijit V, “A simple model of herd behavior,” *The quarterly journal of economics*, 1992, *107* (3), 797–817.

Benabou, Roland and Jean Tirole, “Laws and Norms,” Working Paper, Princeton University 2024.

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

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

— , — , **Omer Tamuz, and Ivo Welch**, "Information cascades and social learning," Technical Report, National Bureau of Economic Research 2021.

Bohren, J Aislinn, "Informational herding with model misspecification," *Journal of Economic Theory*, 2016, *163*, 222–247.

Braghieri, Luca, "Political correctness, social image, and information transmission," *Work. Pap., Stanford Univ., Stanford, CA*, 2021.

Burguet, Roberto and Xavier Vives, "Social learning and costly information acquisition," *Economic theory*, 2000, *15* (1), 185–205.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived social norms: Women working outside the home in Saudi Arabia," *American economic review*, 2020, *110* (10), 2997–3029.

— **and David Y Yang**, "Misperceptions about others," *Annual Review of Economics*, 2022, *14*, 425–452.

— , **Georgy Egorov, and Robert Jensen**, "Cool to be smart or smart to be cool? Understanding peer pressure in education," *The Review of Economic Studies*, 2019, *86* (4), 1487–1526.

— , — , **and Stefano Fiorin**, "From extreme to mainstream: The erosion of social norms," *American economic review*, 2020, *110* (11), 3522–3548.

Canes-Wrone, Brandice, Michael C Herron, and Kenneth W Shotts, "Leadership and pandering: A theory of executive policymaking," *American Journal of Political Science*, 2001, pp. 532–550.

Chandrasekhar, Arun G, Benjamin Golub, and He Yang, "Signaling, shame, and silence in social learning," Technical Report, National Bureau of Economic Research 2018.

Cho, In-Koo and David M Kreps, "Signaling games and stable equilibria," *The Quarterly Journal of Economics*, 1987, *102* (2), 179–221.

Dasaratha, Krishna, Benjamin Golub, and Nir Hak, "Learning from neighbours about a changing state," *Review of Economic Studies*, 2023, *90* (5), 2326–2369.

Ely, Jeffrey C and Juuso Välimäki, "Bad reputation," *The Quarterly Journal of Economics*, 2003, *118* (3), 785–814.

Fernández-Duque, Mauricio, "The probability of pluralistic ignorance," *Journal of Economic Theory*, 2022, *202*, 105449.

Ferreira, Pedro, Selim Gulesci, Eliana La Ferrara, David Smerdon, and Munshi Sulaiman, "Changing Harmful Norms through Information and Coordination: Experimental Evidence from Somalia," Technical Report, mimeo 2024.

Franzen, Axel and Sebastian Mader, "The power of social influence: A replication and extension of the Asch experiment," *Plos one*, 2023, *18* (11), e0294325.

Frick, Mira, Ryota Iijima, and Yuhta Ishii, "Misinterpreting others and the fragility of social learning," *Econometrica*, 2020, *88* (6), 2281–2328.

Geiger, Nathaniel and Janet K Swim, "Climate of silence: Pluralistic ignorance as a barrier to climate change discussion," *Journal of Environmental Psychology*, 2016, *47*, 79–90.

Goeree, Jacob K, Thomas R Palfrey, and Brian W Rogers, "Social learning with private and common values," *Economic theory*, 2006, *28* (2), 245–264.

Golub, Benjamin and Matthew O Jackson, "Naive learning in social networks and the wisdom of crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–149.

Griffin, Kenneth W and Gilbert J Botvin, "Evidence-based interventions for preventing substance use disorders in adolescents," *Child and Adolescent Psychiatric Clinics*, 2010, *19* (3), 505–526.

Kartik, Navin, SangMok Lee, Tianhao Liu, and Daniel Rappoport, "Beyond unbounded beliefs: How preferences and information interplay in social learning," *Econometrica*, 2024, *92* (4), 1033–1062.

Kuran, Timur and William H Sandholm, "Cultural integration and its discontents," *The Review of Economic Studies*, 2008, *75* (1), 201–228.

Lee, In Ho, "On the convergence of informational cascades," *Journal of Economic theory*, 1993, *61* (2), 395–411.

Li, Hongyi and Eric Van den Steen, "Birds of a feather… Enforce social norms? Interactions among culture, norms, and strategy," *Strategy Science*, 2021, *6* (2), 166–189.

Loury, Glenn C, "Self-censorship in public discourse: A theory of "political correctness" and related phenomena," *Rationality and Society*, 1994, *6* (4), 428–461.

Manski, Charles F. and Joram Mayshar, "Private Incentives and Social Interactions: Fertility Puzzles in Israel," *Journal of the European Economic Association*, 3 2003, *1* (1), 181–211.

Michaeli, Moti and Daniel Spiro, "Norm conformity across societies," *Journal of public economics*, 2015, *132*, 51–65.

Miller, Dale T, "A century of pluralistic ignorance: what we have learned about its origins, forms, and consequences," *Frontiers in Social Psychology*, 2023, *1*, 1260896.

_ **and Deborah A Prentice**, "Collective errors and errors about the collective," *Personality and Social Psychology Bulletin*, 1994, *20* (5), 541–550.

Morris, Stephen, "Political correctness," *Journal of political Economy*, 2001, *109* (2), 231–265.

Ottaviani, Marco and Peter Norman Sørensen, "Reputational cheap talk," *The Rand journal of economics*, 2006, *37* (1), 155–175.

Prentice, Deborah A and Dale T Miller, "Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm.," *Journal of personality and social psychology*, 1993, *64* (2), 243.

Scharfstein, David S and Jeremy C Stein, "Herd behavior and investment," *The American economic review*, 1990, pp. 465–479.

Schroeder, Christine M and Deborah A Prentice, "Exposing pluralistic ignorance to reduce alcohol use among college students 1," *Journal of Applied Social Psychology*, 1998, *28* (23), 2150–2180.

Smith, Lones and Peter Sørensen, "Pathological outcomes of observational learning," *Econometrica*, 2000, *68* (2), 371–398.

Sunstein, Cass R, *Going to extremes: How like minds unite and divide*, Oxford University Press, 2009.

Tevyaw, Tracy O'Leary, Brian Borsari, Suzanne M Colby, and Peter M Monti, "Peer enhancement of a brief motivational intervention with mandated college students.," *Psychology of Addictive Behaviors*, 2007, *21* (1), 114.

Tirole, Jean, "Safe spaces: shelters or tribes ?," 2023. Working Paper.

West, Steven L and Keri K O'Neal, "Project DARE outcome effectiveness revisited," *American journal of public health*, 2004, *94* (6), 1027–1029.

Westphal, James D and Michael K Bednar, "Pluralistic ignorance in corporate boards and firms' strategic persistence in response to low firm performance," *Administrative science quarterly*, 2005, *50* (2), 262–298.

8 Appendix A

Proof of Proposition 1. Solving Equation (4) proves the existence of $\kappa^{\text{c.k.}}$ and the results when $\kappa \geq \kappa^{\text{c.k.}}$. Finally, the equilibrium with revelation is Pareto superior to the pooling equilibrium, which completes the proof. □

Proof of Lemma 1. Fix t and a conjectured belief $\alpha_t a_t + \beta_t$. The first-order condition given this conjecture is Equation (9). The normality assumption implies that the posterior expectations of both θ and μ are linear in both v_t and s_t. Therefore, for some exogenous constants, $c_{1,t}, \ldots, c_{6,t}$,

$$a_t(1+\kappa\alpha_t^2) = \kappa\alpha_t\beta_t + c_{1,t} + c_{2,t}\alpha_t\kappa + s_t(c_{3,t} + \kappa\alpha_t c_{4,t}) + v_t(1 + c_{5,t} + \kappa\alpha_t c_{6,t}). \tag{23}$$

$$\iff \frac{a_t(1+\kappa\alpha_t^2) - \kappa\alpha_t\beta_t - c_{1,t} - c_{2,t}\alpha_t\kappa}{c_{3,t} + \kappa\alpha_t c_{4,t}} = s_t + v_t\frac{1 + c_{5,t} + \kappa\alpha_t c_{6,t}}{c_{3,t} + \kappa\alpha_t c_{4,t}}.$$

Given this sufficient statistic, the posterior belief about v_t is,

$$\mathbb{E}\Big(v_t|h_t, a_t\Big) = c_{7,t} + c_{8,t}(\alpha_t,\kappa)\frac{a_t(1+\kappa\alpha_t^2) - \kappa\alpha_t\beta_t - c_{1,t} - c_{2,t}\alpha_t\kappa}{c_{3,t} + \kappa\alpha_t c_{4,t}}, \tag{24}$$

where $c_{8,t}(\alpha_t, \kappa) \neq 0$ is determined by both the prior beliefs and the conjectured slope α_t. In equilibrium, the conjecture is consistent implying the right-hand side of Equation (24) equals $\alpha_t a_t + \beta_t$, as stated below:

$$c_{8,t}(\alpha_t, \kappa)(1 + \kappa\alpha_t^2) = \alpha_t(c_{3,t} + \kappa\alpha c_{4,t}) \qquad (25)$$

$$c_{7,t}(c_{3,t} + \kappa\alpha_t c_{4,t}) + c_{8,t}(\alpha_t, \kappa)(-\kappa\alpha_t\beta_t - c_{1,t} - c_{2,t}\kappa\alpha_t) = \beta_t(c_{3,t} + \kappa\alpha_t c_{4,t}). \qquad (26)$$

Note that for any solution to Equation (25), there exists a solution to the equation for β_t. This is because Equation (26) is linear in β_t with a coefficient on β_t of

$$c_{3,t} + \kappa\alpha_t c_{4,t} + \kappa\alpha_t c_{8,t}(\alpha_t, \kappa) = c_{8,t}(\alpha_t, \kappa)(\frac{1}{\alpha_t} + 2\alpha_t\kappa),$$

where the equality comes from Equation (25). Finally, as $c_{8,t}(\alpha_t, \kappa) \neq 0$ and in a revealing equilibrium $\alpha_t \neq 0$, the coefficient on β_t is non-zero implying a unique solution for β_t.

Thus, the necessary and sufficient condition for an equilibrium with revelation is Equation (25). Further, Equation (25) is independent of the means of the prior beliefs ($c_{1,t}$ and $c_{2,t}$) and only conditions on the precision matrix. With Gaussian uncertainty, the realizations of a_t only effect the mean, proving the second claim.

To show the first claim, I first show $\theta(t) \to_p \theta \implies \mu(t) \to_p \mu$. If $\theta(t) \to_p \theta$, then there exists an infinite subsequence $\{t_i\}_{i=1}^{\infty}$ such that: $c_{3,t_i} + \kappa\alpha_{t_i}c_{4,t_i} \neq 0$, where such terms are defined in Equation (23). However, as $\theta(t) \to_p \theta$, $c_{5,t_i} \to_p 0$ implying $1 + c_{5,t_i} + \kappa\alpha_{t_i}c_{6,t_i} > 1$ in the limit. Therefore Equation (23) implies there are infinitely many noisy signals about v_t, implying $\mu(t) \to \mu$.

I next proceed by contradiction and assume $\mu(t) \to_p \mu$, but $\theta(t) \not\to_p \theta$. Therefore, there exists an infinite subsequence $\{t_i\}_{i=1}^{\infty}$ such that $1 + c_{5,t_i} + \kappa\alpha_{t_i}c_{6,t_i} \neq 0$, but $c_{3,t_i} + \kappa\alpha_{t_i}c_{4,t_i} = 0$. As $c_{3,t_i} + \kappa\alpha_{t_i}c_{4,t_i} = 0$, then $\tilde{s}(\cdot) = 1/(1 + c_{5,t_i} + \kappa\alpha_{t_i}c_{6,t_i})$, implying α_{t_i} must remain bounded to satisfy Equation (11). Finally, given the limit beliefs, $c_{4,t_i} \to_p 0$ and $c_{3,t_i} \not\to_p 0$, contradicting $c_{3,t_i} + \kappa\alpha_{t_i}c_{4,t_i} = 0$, as α_{t_i} is bounded. □

Proof of Lemma 2. I proceed by contradiction. Note that Lemma 1 implies $\theta(t) \to_p \theta \iff \mu(t) \to_p \mu$. Hence, suppose by contradiction that $\kappa > \kappa^{c.k.}$, $\theta(t) \to_p \theta$, and $\mu(t) \to_p \mu$. As $\theta(t) \to_p \theta, \mu(t) \to_p \mu$, then $c_{3,t}, c_{4,t}, c_{5,t}$, and $c_{6,t}$ converge to zero and there exists a subsequence α_{t_i} of revealing decision rules.

Case 1 ($\sup \alpha_{t_i} < \infty$): Equation (23) implies $\lim_{t_i \to \infty} \tilde{s}(\cdot) = 1$. As $\kappa > \kappa^{c.k.}$, no solution exists to Equation (11) asymptotically, implying a switch to a pooling in equilibrium in finite time, contradicting belief convergence.

Case 2 ($\sup \alpha_{t_i} = \infty$): As $c_{6,t} \geq c_{4,t}$ (v_t is more predictive of μ than s_t), then asymptotically the sensitivity is bounded below by

$$\frac{\partial \mathbb{E}(v_t | s_t \kappa \alpha_t c_{6,t} + v_t(1 + \kappa \alpha_t c_{6,t}))}{\partial(s_t \kappa \alpha_t c_{6,t} + v_t(1 + \kappa \alpha_t c_{6,t}))}, \tag{27}$$

which is bounded away from zero. This derives a contradiction to Equation (11) which cannot hold with $\sup \alpha_{t_i} = \infty$ if the sensitivity is bounded away from zero. □

The following remark will be used in subsequent proofs.

Remark 1 *Fix $\theta(t), \mu(t)$, the LHS exceeds the RHS of Equation* (11) *as $\alpha_t \to \infty$.*

Proof of Remark 1. A sufficient condition for the lemma is showing that the sensitivity is bounded as $\alpha_t \to \infty$. Using Equation (23),

$$\lim_{\alpha_t \to \infty} s(\cdot) = \frac{\partial \mathbb{E}(v_t | s_t c_{4,t} + v_t c_{6,t})}{\partial(s_t c_{4,t} + v_t c_{6,t})}, \tag{28}$$

which is bounded. □

Proof of Lemma 3. It suffices to show that the decision rule is revealing in every period; i.e., that there exists a solution to Equation (25). Using Remark 1 and continuity, a sufficient condition is that for all t, there exists an α_t such that

$$c_{8,t}(\alpha_t, \kappa)(1 + \kappa \alpha_t^2) < \alpha_t(c_{3,t} + \kappa \alpha_t c_{4,t}).$$

Since $\kappa < \kappa^{\text{c.k.}}$, then upon setting $\alpha_t = 2$ a sufficient condition is that for some $\epsilon > 0$,

$$\frac{c_{8,t}(2, \kappa)}{c_{3,t} + 2\kappa c_{4,t}} \leq 1 + \epsilon.$$

Further, for $\epsilon > 0$ this inequality holds for sufficiently precise beliefs about θ and μ, because (i) the left-hand side is continuous with respect to the variance matrix of $\theta(t), \mu(t)$, (ii) the left-hand side is equal to one when the variance matrix is equal to zero. Therefore, there exists sufficiently precise prior beliefs where such a condition holds for $t = 1$, and any $t > 1$ as those subsequent beliefs are more precise. □

Proof of Proposition 2. The proof of statement (3) is a direct consequence of Lemma 2 and the proof of statement (2) is a direct result of Lemma 3.

I prove statement (1) by assuming that an equilibrium with revelation does not exist for some $\theta(t), \mu(t)$, and then proving κ must be large. Note that $c_{5,t} \geq -1/2$ (a bound achieved with perfect negative correlation between $\theta(t), \mu(t)$ and arbitrarily diffuse priors). Using Remark 1 and that $\tilde{s}(\cdot) \leq 1/(1 + c_{5,t} + \kappa\alpha c_{6,t})$ this implies

$$\frac{1 + \kappa\alpha_t^2}{1 + c_{5,t} + \kappa\alpha c_{6,t}} > \alpha \text{ for all } \alpha \geq 0 \implies 1 + 2c_{5,t} + c_{5,t}^2 < 4\kappa - 4c_{6,t}\kappa \implies \kappa > \frac{1}{16},$$

where the first implication is the quadratic formula and the second follows from $c_{5,t} \geq -1/2, c_{6,t} \geq 0$, completing the claim.

The final statement is proven using Mathematica to show if $\kappa = .9\kappa^{\text{c.k.}}$ and γ_t is sufficiently close to 1, then there exists an open set of precision matrices for which Equation (25) has no solution for some t.[18] □

Proof of Proposition 3. First, simplify Equation (9) when μ is common knowledge:

$$a_t(1 + \kappa\alpha_t^2) = \frac{\bar{\theta}(t)\tau_{\theta,t} + s_t}{1 + \tau_{\theta,t}} + v_t + \kappa\alpha_t\beta_t. \tag{29}$$

Lemma 1 implies it is without loss of generality to assume $\bar{\theta}(t) = 0$, $\mathbb{E}(s_t) = 0$ and drop the constant $\kappa\alpha_t\beta_t$, for the purposes of calculating the sensitivity. Now,

$$\tilde{s}(\alpha_t, \kappa, \theta(t)) = \frac{\partial \mathbb{E}(v_t | v_t + \frac{s_t}{1+\tau_{\theta,t}})}{\partial(v_t + \frac{s_t}{1+\tau_{\theta,t}})} = \frac{\text{Pre}(\frac{s_t}{1+\tau_{\theta,t}})}{1 + \text{Pre}(\frac{s_t}{1+\tau_{\theta,t}})} := \tilde{\tau}_t(\tau_{\theta,t}), \tag{30}$$

where Pre$(\cdot)$ denotes the precision. Thus, a revealing equilibrium exists if and only if

$$\text{there exists } \alpha_t > 0 \text{ s.t. } (1 + \kappa \cdot \alpha_t^2) \cdot \tilde{\tau}_t(\tau_{\theta,t}) = \alpha_t \iff \kappa \leq \frac{\kappa^{\text{c.k.}}}{\tilde{\tau}_t(\tau_{\theta,t})} \tag{31}$$

If $\kappa \leq \kappa^{\text{c.k.}}$, then this equation holds for all t as $\tilde{\tau}_t(\tau_{\theta,t}) \leq 1$. Further, in every period with revelation $\tau_{\theta,t+1} = \tau_{\theta,t} + 1/(1 + \tau_{\theta,t})^2$, implying $\tau_{\theta,t} \to \infty$ when $\kappa \leq \kappa^{\text{c.k.}}$. Therefore, the comparative statics results follow from the benchmark when $\kappa \leq \kappa^{\text{c.k.}}$.

[18] As discussed in the text a sufficient condition is that the sensitivity is greater than one; however, the sensitivity is endogenous. Therefore, the result is shown using Mathematica, and the code is available upon request.

When $\kappa > \kappa^{\text{c.k.}}$, as $\tilde{\tau}(\tau_{\theta,t})$ is increasing in $\tau_{\theta,t}$ and given the update rule for $\tau_{\theta,t}$ when there is revelation, a higher κ implies weakly less periods with revelation. This statement proves $\tau_\theta(\kappa) := \lim \tau_{\theta,t}(\kappa)$ is decreasing. Further, $\tau_\theta(\kappa) < \infty$ as Equation (31) fails when $\tau_{\theta,t} = \infty$, because $\kappa > \kappa^{\text{c.k.}}$. Finally, when $\kappa > \kappa^{\text{c.k.}}$, the players pool on $\bar{\theta}(t)$, therefore the comparative statics about $\tau_\theta(\kappa)$ imply the comparative statics about asymptotic utility and asymptotic adaptation loss. □

Proof of Proposition 4. The proof of the first statement is a direct consequence of the proof of statement 2 of Proposition 2 and the proof of the second statement follows from statement 1 of Proposition 2.

The proof of the third statement contains two parts. The proof that a pooling equilibrium may be fragile to a peer-oriented intervention is provided in Figure 1 which shows that an increase in $\tau_{\mu,t}$ may break a pool. The proof that an increase in $\tau_{\theta,t}$ never breaks a pooling equilibrium can be seen by noting that when $\theta(t)$ and $\mu(t)$ are independent, Equation (9) reduces to

$$a_t(1+\kappa\alpha^2) = \gamma_t \mathbb{E}\big(\theta \mid \theta(t), s_t\big) + v_t + \kappa\alpha\beta + \kappa\alpha\mathbb{E}\big(\mu \mid \mu(t), v_t\big)$$

Therefore, uncertainty about θ decreases $\tilde{s}(\cdot)$. Further, decreasing $\tilde{s}(\cdot)$ decreases the LHS of Equation (11) and Remark 1 implies the condition for a revealing equilibrium is to find α_t where the LHS is smaller than the RHS of Equation (11). □

Proof of Proposition 5. The players pool on $a^*(a_1) = \mathbb{E}(\theta|a_1) + \mathbb{E}(\mu|a_1)$. Player 1 chooses a_1 as follows:

$$a_1 = \lambda_\theta \mathbb{E}(\theta|s_1) + \lambda_v \mathbb{E}(\mu|v_1) = \lambda_\theta\big(\theta + N(0, 1+\tau_\theta)\big) + \lambda_v\big(\mu + N(0, 1+\tau_\mu)\big).$$

As a result, a^* is linear in a_1. Now consider an individual-oriented intervention where θ is revealed. One must compute $\mathbb{E}(\mu|a_1, \theta)$ by noting:

$$\frac{a_1 - \lambda_\theta\theta}{\lambda_v} = \mu + N(0, 1+\tau_\mu) + \frac{\lambda_\theta}{\lambda_v} N(0, 1+\tau_\theta).$$

As a result,

$$\mathbb{E}(\mu \mid a_1) = \frac{\frac{a_1 - \lambda_\theta \theta}{\lambda_v}\left(\frac{1}{1+\tau_\mu} + (\frac{\lambda_\theta}{\lambda_v})^2 \frac{1}{1+\tau_\theta}\right)^{-1}}{\tau_\mu + \left(\frac{1}{1+\tau_\mu} + (\frac{\lambda_\theta}{\lambda_v})^2 \frac{1}{1+\tau_\theta}\right)^{-1}} = \frac{a_1 - \lambda_\theta \theta}{\lambda_v} \frac{1}{1 + \tau_\mu \left(\frac{1}{1+\tau_\mu} + (\frac{\lambda_\theta}{\lambda_v})^2 \frac{1}{1+\tau_\theta}\right)}.$$

Simplifying this posterior expectation produces the result in the proposition. Further, a symmetric calculation occurs for individual-oriented interventions. □

Proof of Proposition 6. First note that Proposition 7 rules out non-linear equilibria when $\kappa > \kappa^{\text{c.k.}} = 1/4$ as $\xi_t = 0$ in this benchmark. Hence, let $\kappa \leq \kappa^{\text{c.k.}}$.

As mentioned in the text, there always exists the fully pooling decision rule where $-\underline{v} = \bar{v} = \infty$. Hence, let us consider central pooling equilibria that involve revelation, and I will show that outside the central pool the decision rule is linear. As the problem is symmetric about $v = 0$, it is without loss to assume $\bar{v} < \infty$. First, if $a(v)$ were bounded then Equation (19) implies $\lim_{v \to \infty} a'(v) = \kappa$ contradicting $a(v)$ being bounded. Therefore, $\lim_{v \to \infty} a(v) = \infty$.

The first-order condition implies:

$$a - v + \kappa \hat{v}(a) \hat{v}'(a) = 0.$$

Substituting $a = x$ and $ay(x) = \hat{v}(x)$ above yields,

$$0 = x - xy(x) + \kappa x y(x)\big(y(x) + x\dot{y}(x)\big) \iff \dot{y}(x) = \frac{-1 + y(x) - \kappa y(x)^2}{\kappa x y(x)}. \qquad (32)$$

This simplification is well defined because the denominator is non-zero for all x outside the central pool. If $\dot{y}(x_0) = 0$, then $y(x_0)$ agrees with one of the linear decision rules, $l(\cdot)$ (which exist as $\kappa \leq \kappa^{\text{c.k.}}$). Therefore (i) $A := \{x : y(\cdot) = l(\cdot)\}$ is non-empty and closed. As we are assuming the decision rule is non-linear then, (ii) $A \neq [a(\bar{v}), \infty]$. Combining (i) and (ii), there exists a point in A on the boundary of A. However, at such a point, the above differential equation satisfies the conditions for Picard–Lindelöf theorem, implying a unique solution within an interval surrounding that point, contradicting that such a point lies on the boundary.

Therefore any non-linear solution satisfies $\dot{y}(x) \neq 0 \forall x$, and for any $x > 0$, one can derive the following from Equation (32):

$$\int \frac{\dot{y}(x)\kappa y(x)}{-1+y(x)-\kappa y(x)^2} = \int \frac{-1}{x} \iff \tag{33}$$

$$\frac{-\log(1-y(x)+\kappa y(x)^2)}{2} + \frac{\log(1+\frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}) - \log(1-\frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}})}{2\sqrt{1-4\kappa}} + c = -\log(x),$$

where c is the constant of integration. If $\kappa = 1/4$, then Equation (33) is ill-defined, precluding a non-linear solution. If $\kappa < 1/4$, any non-linear solution satisfies:

$$\log\left(\sqrt{1-y(x)+\kappa y(x)^2} \cdot \left(\frac{1-\frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}{1+\frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}\right)^{\frac{1}{2\sqrt{1-4\kappa}}}\right) = \log(e^c x)$$

$$\iff (1-y(x)+\kappa y(x)^2)\left(\frac{1-\frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}{1+\frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}\right)^{\frac{1}{\sqrt{1-4\kappa}}} = x^2 e^c.$$

Recall that $\lim_{v\to\infty} a(v) = \infty$ and thus one can consider the limit of the above equation as $x \to \infty$. The RHS diverges as $x \to \infty$ implying the LHS diverges. Further, if $\lim_{x\to\infty} y(x) < \infty$, then both the first and second terms of the left-hand side will remain bounded. Therefore, $\lim_{x\to\infty} y(x) = \lim_{v\to\infty} v/a(v)\infty$. However, if $\lim_{v\to\infty} v/a(v) = \infty$, then there must exist a player for which ϵ less conformity is preferred, deriving a contradiction.

Therefore, the only solutions to the differential equation outside the central pool are linear. I will now show that the central pool is empty. Equation (4) shows that the linear equilibria must satisfy:

$$\alpha(1-\alpha) = \kappa \iff \alpha = \frac{1 \pm \sqrt{1-4\kappa}}{2}.$$

Thus for a central pooling equilibrium to exist one must find $a^*, \underline{v}, \bar{v}$ such that the player with preference type $\bar{v}$ is indifferent between a^* and $\alpha\bar{v}$. I consider the case where $a^* \leq 0$ and an identical argument holds if $a^* \geq 0$.[19] Note that the conformity loss following a^* is at best zero. Fixing the slope of the linear decision rule $a(v) = \alpha v$,

[19] One can use $\underline{v}$ analogously to $\bar{v}$. If $\underline{v}$ was negative infinity, then the equilibrium could not satisfy D1 as no beliefs satisfying D1 following $a^* - \epsilon$ would prevent a player with an arbitrarily negative preference type from deviating.

then,

$$-(\bar{v}-a^*)^2 \geq -(1-\alpha)^2\bar{v}^2 - \kappa\bar{v}^2,$$

where the LHS is an upper bound on the utility in the central pool and the RHS is the utility in the linear equilibrium. Notice that $\bar{v}^2$ cancels out and that for the solutions to α such an inequality never holds. Therefore, the central pool must be empty in any revealing equilibrium. □

Proof of Proposition 7. First note that there exists a unique v_t^* such that $v_t^* = g_t(v_t^*)$, where uniqueness follows from the assumption that $g_t'(\cdot) < 1$ and existence follows from the Law of Iterated expectations and continuity of $g_t'(\cdot)$. By an identical argument to Theorem 3 in Bernheim (1994) all equilibria must be central pooling, the central pool must include v_t^*, and Equation (20) characterizes the equilibrium outside of the central pool.

I first show that if κ exceeds the threshold in the proposition no solution to Equation (20) exists for any initial condition, implying no equilibrium with revelation exists. Differentiating Equation (20) implies,

$$\begin{aligned} a_t''(v_t) &= \frac{-\kappa\big(v_t - g_t(v_t)\big)\big(1-a_t'(v_t)\big)}{\big(v_t - a_t(v_t)\big)^2} + \frac{\kappa\big(1-g_t(v_t)\big)}{v_t - a_t(v_t)} \\ \iff \big(v_t - g_t(v_t)\big)a_t''(v_t) &= a_t'(v_t)\left(1 - g_t'(v_t) - \frac{a_t'(v_t)\big(1-a_t'(v_t)\big)}{\kappa}\right). \end{aligned}$$

I will analyze values to the right of the central pool (and a symmetric argument follows for the left side). As $a_t'(v_t)\big(1-a_t'(v_t)\big) \leq 1/4$, then for $\kappa > 1/4(1-\xi_t)$ there exists an $\epsilon_3 > 0$ such that,

$$\big(v_t - g_t(v_t)\big)a_t''(v_t) \geq a_t'(v_t)\epsilon_3 \iff a_t''(v_t) \geq \frac{a_t'(v_t)\epsilon_3}{v_t - g_t(v_t^*)}, \tag{34}$$

where the final inequality comes from the assumed monotonicity of $g_t(\cdot)$. Let $a_0(v_t)$ be the decision rule that binds differential inequality in Equation (34) and satisfies $a_t(\bar{v}_t) = a_0(\bar{v}_t)$, where $\bar{v}_t$ denotes the supremum of the central pool.

One can use the Picard-Lindelöf Theorem to show $a_0(v_t)$ has a unique solution up to this initial condition satisfying

$$a_0(v_t) = c_1(v_t - g_t(v_t^*))^{1+\epsilon_3} + c_2,$$

where $c_1 > 0$. Therefore for $v_t > v_t^*$,

$$a_t(v_t) = a_t(\bar{v}_t) + \int\int a_t''(v_t) \geq a_0(\bar{v}_t) + \int\int a_0''(v_t).$$

However, given the solution for $a_0(\cdot)$, this inequality implies that $a_t(v_t) > v_t$ for a positive value of v_t, which is a contradiction because choosing $a_t = v_t$ results in both a better adaptation loss and a better conformity loss.

Now I show that if κ is less than the condition provided in Proposition 7 an equilibrium with full revelation exists. To do so, I show that if κ satisfies the condition in the proposition there exists an equilibrium with full revelation where $a_t(v_t^*) = v_t^*$ by using Carathéodory's existence theorem.

To be able to apply this theorem to Equation (20) the decision rule, $a_t(v_t)$ must never cross v_t except for v_t^* so that the implicit function in Equation (20) is continuous on its domain. I will analyze the differential equation to the right of v_t^* and a symmetric analysis occurs to the left. A sufficient condition for $a_t(v_t) < v_t$ is that $a_t'(v_t) \leq 1/2$ for all $v_t \geq v_t^*$. Taking Equation (20), this condition can be stated as

$$2\kappa\big(v_t - g_t(v_t)\big) \leq \frac{v_t - a_t(v_t)}{2} \impliedby \kappa\big(v_t - g_t(v_t)\big) \leq \frac{v_t - v_t^*}{4} \iff \kappa \leq \frac{\kappa^{\text{c.k.}}}{1 - \frac{g_t(v_t) - v_t^*}{v_t - v_t^*}}.$$

Applying the following bound to the denominator completes the proof:

$$\frac{g_t(v_t) - v_t^*}{v_t - v_t^*} \geq \frac{g_t(v_t^*) + \iota_t(v_t - v_t^*) - v_t^*}{v_t - v_t^*} = \iota_t,$$

as $g_t(v_t^*) = v_t^*$. □