Conformity Concerns: A Dynamic Perspective

Roi Orzach*

February 17, 2026

Abstract

In many settings, individuals imitate their peers' decisions both to adapt to a common fundamental state and to conform to their peers' preferences. I study a dynamic model where both the fundamental state and peers' preferences are initially unknown, and players learn by observing others' decisions. As public beliefs become more precise, conformity motives strengthen endogenously, eventually resulting in permanently uninformative decision-making. Therefore, when conformity concerns are large, persistent misperceptions about both objects arise. Consistent with evidence from social psychology, I identify environments where correcting misperceptions of the fundamental state is less effective than correcting misperceptions of peers' preferences.

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, Krishna Dasaratha, 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, Juan Ortner, Shelli Orzach, Jacopo Perego, Nicola Persico, Kramer Quist, Daniel Rappoport, Evan Sadler, Rafaella Sadun, Tomasz Sadzik, Lones Smith, Jean Tirole, Jonathan Weinstein, Alexander Wolitzky, as well as participants at North American Summer Meeting of the Econometric Society, Stony Brook Game Theory Conference, Theoretical Research in Development Economics conference and seminar participants at the MIT Organizational Economics and Theory groups, Tel Aviv University, Washington University, Northwestern, Stony Brook, Boston University, University of Chicago, University of Georgia, University of Haifa, and Hebrew University 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, including political speech (Loury, 1994), female labor force participation (Bursztyn et al., 2020a), and medical behaviors (Ferreira et al., 2024). Table 1 summarizes how these settings map into the model. 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 the group's 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' preferences because they are drawn from the same population distribution. Socially, each player observes their peers' decisions, which may be informative about their signals or preferences. 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 maximizes a utility function with two parts: a private payoff and a conformity payoff. The private payoff depends on how well the chosen action adapts to the individual's own preferences and the fundamental state. The conformity payoff equals a common and exogenous conformity parameter multiplied by the gap between the community's perception of the individual's preferences and the community's average preferences. I show that if social learning fails, then asymptotically the players face three inefficiencies.

First, the players are unable to adapt to their preferences. Further, the players' decisions are based on imprecise beliefs about *both* the fundamental state and the community's average preferences.

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 outlined in Table 1: social learning fails and effective informational interventions are about peers' preference types.

Main Results For most of the analysis, I assume Gaussian uncertainty and a linearity refinement; these assumptions are relaxed in Section 7.

In Section 3, I first analyze a benchmark static model where players have common knowledge about both the fundamental state and their peers' average preferences but differ in their own preferences, as described above. Here, conformity concerns place a penalty on the degree to which players adapt to their preferences (cf. Bernheim, 1994). I show that all players choose the same decision independent of their preferences 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 preferences because doing so would incur a large reputational penalty. Further, because adaptation to a player's private signal about the fundamental state would be attributed to a player's preference type, such adaptations are similarly discouraged. Therefore, when the conformity concerns are large enough, each decision is uninformative about a player's private information. Further, Proposition 2 shows that social learning about preferences (respectively, fundamentals) occurs if and only if the conformity concerns are small.

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]

[1] I discuss the difference between my benchmark and the environment in Bernheim (1994) in Section 7.1. Reassuringly, in Bernheim (1994) when conformity concerns are sufficiently high, the unique equilibrium is also fully pooling.

[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

Absent a desire for conformity, individuals would disregard peer-oriented interventions entirely. Instead, my analysis shows that (i) only peer-oriented interventions can break pooling equilibria (Proposition 3), and (ii) individual-oriented interventions may have no effect on which decision the players pool on (Proposition 4).

Section 6 relates these theoretical findings to several influential empirical results. First, Braghieri (2021) provides experimental evidence that social-image concerns distort public speech on sensitive sociopolitical issues; accordingly, speech is informative about underlying preferences if and only if social-image concerns are low. Interpreting actions as stated opinions, this evidence aligns with the model's prediction that conformity concerns can render behavior uninformative about private information. In addition, the paper's implications for how inefficiencies can be mitigated connect to a growing empirical literature on informational interventions. For instance, 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. Finally, similar qualitative patterns are documented in Ferreira et al. (2024) and Pei et al. (2025).

Section 7 relaxes the assumptions of Gaussian uncertainty and linearity while providing additional comparative statics. Section 8 concludes and the Appendix provides proofs for all claims in the text.

Related Literature This paper is related to two strands of theoretical literature: social learning and decision-making with reputational concerns.

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

differing preferences from the community, resulting in a commitment problem.

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

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 Holmström (1999) or Morris (2001), assumed the definition of a "good type" was common knowledge.[4] In contrast to my setting, in these cases even if the players place arbitrarily high weight on reputation, equilibrium decisions remain informative about a player's type when the action space is unbounded (cf. Holmström, 1999).

In my model, 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 concerns 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). 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 where 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.

[4]Many related papers exist which also assume the "good type" is common knowledge: Scharfstein and Stein (1990); Canes-Wrone et al. (2001); Ely and Välimäki (2003); Ottaviani and Sørensen (2006).

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,

$$\begin{matrix}\theta \\ \mu\end{matrix}\Bigg) \sim \mathcal{N} \quad \begin{matrix}0 \\ 0\end{matrix}\Bigg)\Bigg(\begin{matrix}\sigma_\theta^2 & \rho\sigma_\theta\sigma_\mu \\ \rho\sigma_\theta\sigma_\mu & \sigma_\mu^2\end{matrix}\Bigg)\Bigg)\Bigg(\tag{1}$$

where $\sigma_\theta, \sigma_\mu \geq 0$ and $\rho \in (-1, 1)$.[5] Each player's private information is his signal $s_t = \theta + \epsilon_t$ 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 variances 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.[6] Let $\phi(b|h^t, a_t)$ be the probability distribution over preference types b that take decision a_t given the public history, h^t, in equilibrium. Further, for now, $\phi(\cdot)$ is unconstrained off-path. The total utility for player t is thus,

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

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 parentheses 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. I study Perfect Bayesian Equilibria of this game subject to the following assumptions and refinements.

[5] Section 7 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 9). Finally, to rationalize the empirical findings in Schroeder and Prentice (1998) comparing interventions providing information about θ to those about μ discussed in the introduction, I allow for uncertainty about θ and μ.

[6] 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 $c \neq 0$.

2.1 Assumptions, Refinements, and Examples:

I begin by discussing two assumptions: the modeling of conformity and the heterogeneity assumption. Next, I discuss the equilibrium refinements.

Modeling of Conformity: As in Bernheim (1994), I model conformity (a desire to be perceived as similar to one's peers) as opposed to coordination (a desire to choose similar actions to one's peers). I note that upon replacing the conformity term with a coordination term, such as $\kappa(d_t - \frac{1}{T}\sum_{t' \in T} d_{t'})^2$, there would be efficient learning for any κ. Finally, as in Bernheim (1994), I assume that a player's optimal perception is μ. Given that the loss function is quadratic, one can show that from the standpoint 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.

Observed Heterogeneity: I assume that the public history is $h^t := \{\gamma_1, a_1, \ldots, \gamma_{t-1}, a_{t-1}, \gamma_t\}$. One can interpret γ_t as a commonly observed time or individual fixed effect that determines how important the fundamental state is relative to private preferences. Formally, I assume γ_t is i.i.d. over time with full support on $[0, 1]$. Intuitively, this assumption implies that players place different weights on the fundamental state at different times (e.g., after finals in the drinking example or during election cycles in political speech examples). This heterogeneity is only needed when players must learn both θ and μ: with a one-dimensional action and constant γ, the history generates too few independent restrictions to separately identify two latent objects. Many social-learning models avoid this issue by allowing a richer action space (e.g., multi-dimensional actions), but to keep actions one-dimensional in line with the motivating examples, I instead introduce variation in γ_t. Other sources of variation would suffice as well; for example, time-variation in the signal structure s_t.

Given the equilibrium multiplicity inherent in signaling games, I employ the following refinements to select a unique equilibrium and to produce precise comparative statics.

Definition 1 *A signaling equilibrium is a Perfect Bayesian Equilibrium satisfying the following refinements in each period.*

1. *Linearity: If at h^t $\text{Var}(a_t|h^t) \neq 0$, then $\int b \cdot \phi(b|h^t, a_t)db$ is linear and increasing in a_t.*
2. *Social Optimality: Denote by $\pi^t(h^t)$ the player's beliefs about θ and μ at history h^t, and let $\mathcal{E}(\pi^t)$ be the set of linear equilibria that induce beliefs π^t in period t. An equilibrium, E^*, satisfies social optimality if for all histories $E^* \in \arg\max_{E' \in \mathcal{E}(\pi^t)} \mathbb{E}(u_t(\cdot)|E', \pi^t(h^t))$.*

Linearity: This refinement states that if at h^t the decision rule is not fully pooling, then the expected beliefs about a player's type are linear in their action. As in Ball (2019), this

refinement allows for greater tractability and demands less sophistication from the players—running regressions is sufficient. Further, as players can deviate to any action, linearity is a refinement rather than a restriction. Finally, in Section 7, I discuss how these results extend beyond the Linear-Gaussian environment.

Social Optimality: It is useful to describe the refinement informally. Here, period 1 selects the linear equilibrium that maximizes player 1's ex ante payoff (given h^1); after the induced play updates the public history to h^2, period 2 selects the payoff-maximizing continuation equilibrium for player 2, and so on. If at h^t only pooling linear equilibria exist, the refinement selects the pooling equilibrium that maximizes the current player's payoff; in that case the choice is payoff-irrelevant for other players. Moreover, whenever a revealing linear equilibrium exists at h^t, the refinement selects a revealing one.[7] Without an optimality refinement, pathological equilibria exist: for any sequence $\{x_t\}_{t=1}^{\infty}$, there is an equilibrium in which players pool on x_t in period t.

3 Common Knowledge Benchmark

This section analyzes the impact of conformity on decision-making and eliminates uncertainty about the fundamental state and the preferences of others. To do so, I assume θ and μ are common knowledge, and given that they are common knowledge, it is without loss of generality to assume both are equal to zero.[8] Further, without uncertainty there is no time dependence, so 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 or a strictly increasing decision rule. If the decision rule is constant (respectively, strictly increasing), then it is called "fully-pooling" (respectively, "revealing"). The following proposition characterizes the signaling equilibrium.

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

$$a(v) = \begin{cases} \frac{1+\sqrt{1-4\kappa}}{2}\, v & \text{if } \kappa \leq \kappa^{c.k.}, \\ 0 & \text{if } \kappa > \kappa^{c.k.}, \end{cases} \qquad \text{and} \qquad u(v) = \begin{cases} -\frac{1-\sqrt{1-4\kappa}}{2}\, v^2 & \text{if } \kappa \leq \kappa^{c.k.}, \\ -v^2 - \kappa & \text{if } \kappa > \kappa^{c.k.}. \end{cases} \tag{3}$$

[7] Generically there are two revealing equilibria. The refinement also selects the stable equilibrium in the sense of Echenique (2002). In the one-dimensional case, it coincides with maximizing discounted expected utility over the class of equilibria that are linear in every period.

[8] As θ is common knowledge, players disregard signals s_t. This benchmark is similar to the analysis in Bernheim (1994). Section 7.1 examines this benchmark without assuming linearity or Gaussianity and offers a comprehensive comparison with Bernheim (1994).

When $\kappa = 0$, conformity is irrelevant and each individual chooses $a(v) = v$. As κ increases, individuals place more weight on appearing average and therefore choose actions closer to the mean action, $a = 0$. In the revealing equilibrium, this implies a smaller slope of the decision rule. In any revealing equilibrium, actions fully identify types on path, regardless of how small the slope is. Therefore, when κ is large, deviating to the mean action $a = 0$ yields a large reputational gain as an individual is now perceived as $v = 0$ (the ideal perception). Further, the adaptation cost of such a deviation is small because equilibrium actions are already close to zero, due to the flatter decision rules. Hence, for sufficiently high κ, the signaling equilibrium is fully pooling. Importantly, a fully-pooling equilibrium reveals no information about a player's private information, which will be key in the main analysis.

4 Analysis

This section begins by defining notation and then characterizes learning outcomes.

4.1 Preliminaries

Define $\theta(h^t) = \mathbb{E}(\theta|h^t)$ and $\mu(h^t) = \mathbb{E}(\mu|h^t)$.[9] These random variables have the following joint distribution:

$$\left.\begin{matrix}\theta(h^t)\\ \mu(h^t)\end{matrix}\right) \sim \mathcal{N} \quad \left.\begin{matrix}\bar{\theta}(t)\\ \bar{\mu}(t)\end{matrix}\right)\left(\begin{matrix}\sigma^2_{\theta,t} & \rho_t\sigma_{\theta,t}\sigma_{\mu,t}\\ \rho_t\sigma_{\theta,t}\sigma_{\mu,t} & \sigma^2_{\mu,t}\end{matrix}\right)\Bigg)\Bigg(\tag{4}$$

Definition 2 (**Social Learning**)
Social learning about fundamentals occurs if and only if for all realizations of θ, $\theta(h^t) \to_p \theta$ *and fails when for all realizations of* θ, $\theta(h^t) \not\to_p \theta$. *Social learning about preferences occurs if and only if for all realizations of* μ, $\mu(h^t) \to_p \mu$ *and fails when for all realizations of* μ, $\mu(h^t) \not\to_p \mu$.

This definition permits the possibility that social learning about preferences (or fundamentals) neither fails nor occurs; the analysis shows this never occurs.[10]

In the signaling equilibrium, the inference function, $\phi(b|h^t, a_t)$, is Gaussian, the mean of which I denote by $\hat{v}(a_t) := \int b \cdot \phi(b|h^t, a_t)db$. Further, the variance in the posterior beliefs

[9] Even if I restricted attention to the case where $\theta(h^1)$ and $\mu(h^1)$ are independent, in subsequent periods, $\theta(h^t)$ and $\mu(h^t)$ are dependent as both condition on h^t. Further, at any h^t, $\theta(h^t), \mu(h^t)$ are sufficient statistics for the probability distribution determining the players' beliefs. Therefore, I refer to these random variables as "the beliefs."

[10] I do not characterize the speed of learning, although when learning occurs it is quite fast. Intuitively, learning requires infinitely many revealing periods, and with a linear decision rule each such period yields an observable one-dimensional statistic of the player's private information.

about a player's type, $\int (b - \hat{v}(a_t))^2 \cdot \phi(b|h^t, a_t)db$, is independent of a_t, allowing for the following simplification of the conformity loss up to some constant c_t as stated below:

$$\int (b - \mu)^2 \phi(b|h^t, a_t)db = (\hat{v}(a_t) - \mu)^2 + c_t. \tag{5}$$

Finally, I define the following notions of efficiency.

Definition 3 (**Asymptotic Utility and Adaptation Loss**)
The asymptotic utility *is* $\lim_{t\to\infty} \mathbb{E}\big(u_t(a_t; v_t, s_t|\theta(h^t), \mu(h^t), \gamma_t)\big)$. *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 Expectations.

4.2 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 $\theta(h^t), \mu(h^t)$ are sufficient statistics for the past history.

I first analyze when a revealing equilibrium exists. In any linear revealing equilibrium $\hat{v}(a_t) = \alpha_t a_t + \beta_t$, for some conjectured coefficients α_t, β_t. I now derive player t's type-dependent best-reply to this conjecture through a first-order condition of a player's utility given the simplification in Equation (5):

$$a_t(1 + \kappa\alpha_t^2) = \gamma_t \mathbb{E}\big(\theta \mid \theta(h^t), \mu(h^t), s_t, v_t\big) + v_t - \kappa\alpha_t\beta_t + \kappa\alpha_t \mathbb{E}\big(\mu \mid \theta(h^t), \mu(h^t), s_t, v_t\big). \tag{6}$$

Further, a revealing equilibrium requires that the posterior mean of v_t induced by the decision rule in (6) coincides with the conjectured mapping $\hat{v}(a_t)$ on path. Given the Gaussian uncertainty and the expression for a_t in Equation (6), the posterior expectation of v_t given a_t has the following form:

$$\mathbb{E}\big(v_t \mid a_t, \theta(h^t), \mu(h^t)\big) := a_t(1 + \kappa\alpha_t^2) \cdot \tilde{s}\big(\alpha_t, \kappa, \theta(h^t), \mu(h^t)\big) + \iota\big(\alpha_t, \kappa, \beta_t, \theta(h^t), \mu(h^t)\big), \tag{7}$$

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. Therefore, consistency of the beliefs requires that there exists

an α_t and β_t which satisfy,

$$(1+\kappa\alpha_t^2)\tilde{s}\big(\alpha_t,\kappa,\theta(h^t),\mu(h^t)\big) = \alpha_t \tag{8}$$

$$\iota\big(\alpha_t,\kappa,\beta_t,\theta(h^t),\mu(h^t)\big) = \beta_t. \tag{9}$$

One can see from inspecting Equation (8) that increasing the sensitivity has a similar effect to increasing the conformity concerns. The intuition is that if the sensitivity is larger, then the player's reputation is more sensitive to their decision, increasing their desire to conform. The sensitivity is defined so that when θ, μ are common knowledge, $\tilde{s}(\cdot) = 1$, since absent uncertainty, variation in a player's decision is only attributed to v_t. A solution to Equation (8) is necessary for a revealing equilibrium to exist. While Equation (8) imposes only consistency in the slope of the expectation of the beliefs, the following lemma shows that it is also sufficient.

Lemma 1 ***(Characterization of Revealing Equilibria)***
The signaling equilibrium is revealing in period t if and only if there exists an $\alpha_t \geq 0$ satisfying Equation (8)*.*

Recall that the sensitivity is defined so that, if θ, μ are common knowledge, the sensitivity equals one. Therefore, if $\kappa > \kappa^{\text{c.k.}}$ (i.e., there is no revealing equilibrium if θ, μ are common knowledge), then the sensitivity cannot converge to one while maintaining an equilibrium with revelation. Further, if $\theta(h^t), \mu(h^t)$ are sufficiently precise, then player t effectively disregards s_t and v_t when computing his posterior expectations of θ and μ. Therefore, the right-hand side of Equation (6) is approximately equal to v_t, which implies that the sensitivity is approximately equal to 1, as in the benchmark. This can be used to show that if $\kappa > \kappa^{\text{c.k.}}$, social learning about either fundamentals or preferences must fail. The subsequent lemma shows that social learning fails for both fundamentals and preferences.

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

While the above intuition implies that social learning about both dimensions cannot occur, one might conjecture that there exists an equilibrium where social learning about θ occurs but social learning about μ fails. This intuition would posit that because players have conformity concerns about v_t, not s_t, there may exist an equilibrium where players' decisions adapt to s_t but not v_t. However, this would then imply that there is no reputational penalty from adaptation, thus implying players do adapt their decision to v_t, deriving a contradiction.

The correct intuition is that if, by contradiction, social learning about either fundamentals or preferences occurs, then there must exist infinitely many periods where the signaling

equilibrium is revealing. Given the first-order condition in Equation (6), the decision must place positive weight on both s_t and v_t, implying that the players must learn both θ, μ. Therefore, asymptotically, the environment converges to the environment in the benchmark. However, as previously argued, this cannot occur if $\kappa > \kappa^{\text{c.k.}}$. As a consequence, there must exist a finite time after which the players are pooling in every period. The following proposition nests this result and provides a converse.

Proposition 2 (**Characterization of Social Learning**)
If $\kappa \leq \kappa^{c.k.}$, then social learning about preferences and fundamentals occurs. In contrast, if $\kappa > \kappa^{c.k.}$, then social learning about preferences and fundamentals fails.

Below, I provide intuition for why if $\kappa \leq \kappa^{\text{c.k.}}$ and the beliefs about θ, μ are independent, then the equilibrium is revealing. The formal proof shows that the equilibrium is also revealing with correlation, which may arise in equilibrium as, for instance, high decisions could arise due to either a high θ or a high μ. Further, the proof shows that infinitely many periods of revelation combined with the heterogeneity from γ_t is sufficient for social learning about preferences and fundamentals.[11]

To show the equilibrium is revealing when $\kappa \leq \kappa^{\text{c.k.}}$ and the beliefs are independent, it suffices to show that the endogenous reputational penalty from adaptation is less than that in the benchmark (i.e., $\tilde{s}(\cdot) \leq 1$). This is sufficient because if the equilibrium was revealing in the benchmark (i.e., $\kappa \leq \kappa^{\text{c.k.}}$), then it will continue to be revealing with uncertainty, as the penalty from adaptation is weakly lower.

To show that the reputational penalty is less than that of the benchmark, note that the first-order condition in Equation (6) can be written (up to an additive constant) as $\beta_s s_t + \beta_v v_t$. Under the independence assumption, two properties hold: (i) $\beta_s > 0$ and (ii) $\beta_v > 1$. For (i), s_t affects the first-order condition only through $\mathbb{E}(\theta|h^t, s_t, v_t)$, which is increasing in s_t, implying the players positively weight s_t in their decision. For (ii), v_t affects the first-order condition through $v_t + \kappa\alpha\mathbb{E}(\mu|h^t, s_t, v_t)$; since $\mathbb{E}(\mu|h^t, s_t, v_t)$ is increasing in v_t, the overall coefficient on v_t in the decision rule exceeds one. Together, these properties imply a lesser reputational penalty than the benchmark. First, given (i) (i.e., $\beta_s > 0$), variation in a_t is partly attributed to variation in s_t rather than solely to v_t, thereby decreasing the reputational cost of adaptation. Further, through (ii) (i.e., $\beta_v > 1$), an ϵ increase in a_t corresponds to an $\epsilon/\beta_v < 1$ increase in v_t. Consequently, relative to the benchmark (where

[11] Recall that this heterogeneity rules out a degenerate (and qualitatively uninteresting) scenario in which the equilibrium is revealing in every period but the public repeatedly observes the same fixed linear combination of signals (for example, a decision of the form $\frac{1}{2}s_t + \frac{3}{2}v_t$ each period) which would prevent separately identifying θ and μ. In such a case, the beliefs converge to an unstable learning outcome with perfectly negatively correlated beliefs, as discussed in Online Appendix Section 1.4.

$\beta_v = 1$), a given change in a_t induces a smaller revision in beliefs about v_t, which also implies that the reputational penalty of adaptation is less than that of the benchmark.

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.[12] I discuss this finding in the context of the empirical literature in Section 6.1.

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.

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 the community is not aware that the player received such information 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 or to influence which decision the players pool on.[13] 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 observed s_t such that $\mathbb{E}(\theta|h^t, s_t) = \theta$. Further, since such a signal 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 pooling. 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.

[12] At this level of generality, continuous learning outcomes, such as the asymptotic variance in the beliefs, are non-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, resulting in different beliefs in the subsequent period, thus changing whether the next period is pooling or not. In either analysis with only one dimension of uncertainty, the learning outcomes are strictly monotone with respect to κ, as shown in the next subsection.

[13] 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 for private interventions.

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 $h^t = \{\gamma_1, a_1, \ldots, \gamma_{t-1}, a_{t-1}, \gamma_t\}$, 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, which extends the definition of fragility in Bikhchandani et al. (2021) to signals about μ.

Definition 4 (**Fragility Definition**) *If the signaling equilibrium is pooling in period t, then*

1. *The equilibrium is fragile to an individual-oriented intervention with n signals if a hypothetical public disclosure of n independent signals, each distributed identically to s_t, would make the equilibrium in period t revealing.*

2. *The equilibrium is fragile to a peer-oriented intervention with n signals if a hypothetical public disclosure of n independent signals, each distributed identically to v_t, would make the equilibrium in period t revealing.*

Proposition 3 (**Fragility**)
The following are true when $\kappa > \kappa^{c.k.}$:

1. *For any $n \in \mathbf{N}$, an individual-oriented intervention (respectively, peer-oriented intervention) with n pieces of information will never result in social learning about μ (respectively, θ).*

2. *If the correlation between the beliefs about θ and μ is equal to zero, then 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.*

As social learning about preferences and fundamentals occurs when $\kappa \leq \kappa^{\text{c.k.}}$, the results focus on the case where $\kappa > \kappa^{\text{c.k.}}$. Result 1 says that giving information about only one dimension of uncertainty will be unsuccessful in spurring social learning on the other. Result 1 follows directly from Proposition 2, which states that regardless of the initial beliefs, if $\kappa > \kappa^{\text{c.k.}}$, the players' beliefs about θ and μ cannot converge to the truth. Therefore, for any

individual-oriented intervention, the players will necessarily continue to have misperceptions about μ and vice versa.

An immediate corollary of Result 1 is that neither type of intervention can result in a permanent change to a revealing equilibrium. Result 2 describes when such interventions can cause temporary changes to a revealing equilibrium, as encapsulated in the definition of fragility. Before providing the intuition for Result 2, I note that this result is stated for the case of zero correlation in the beliefs. Therefore, Result 2 applies to the case of an intervention in period 1 or an intervention in a period when every prior period had been pooling. Restricting attention to this case allows for greater analytical tractability and also the ability to isolate the direct informational effects as opposed to effects through correlation. Finally, by continuity, these results will continue to hold for small amounts of correlation.

The intuition for Result 2 comes from the differential effects of these interventions on the sensitivity. A decision rule is fragile when decreasing the variance in the beliefs (as a result of the intervention) decreases the sensitivity. This decrease in the sensitivity results in a lesser reputational penalty for adapting to a player's private information, making it easier to sustain an equilibrium with revelation. Therefore, a pooling decision rule is fragile to an individual-oriented intervention (respectively, peer-oriented) if the sensitivity is increasing in the variance of the beliefs about θ (respectively, μ). The following lemma characterizes the comparative statics of the sensitivity with respect to the variance in the beliefs.

Lemma 3 (**Differential Effects of Interventions on Sensitivity**)
Suppose the correlation between the beliefs about θ and μ is equal to zero. The sensitivity, defined in Equation (7)*, is decreasing in the variance of the beliefs about θ, $\sigma^2_{\theta,t}$. In contrast, the sensitivity is increasing in the variance of the beliefs about μ, $\sigma^2_{\mu,t}$, if and only if*

$$\frac{\gamma_t^2 \sigma^4_{\theta,t}}{1+\sigma^2_{\theta,t}} \geq \frac{\alpha\kappa(1+\sigma^2_{\mu,t}+\alpha\kappa\sigma^2_{\mu,t})^2}{(1+\alpha\kappa)(1+\sigma^2_{\mu,t})^2}, \tag{10}$$

where the right-hand side is strictly positive and increasing in $\sigma^2_{\mu,t}$.

To gain intuition behind these comparative statics, note that if there is no correlation between the beliefs about θ and μ, then the player's first-order condition in Equation (6) is

$$a_t(1+\kappa\alpha^2) = \gamma_t \mathbb{E}(\theta|\theta(h^t), s_t) + v_t + \kappa\alpha\mathbb{E}(\mu|\mu(h^t), v_t). \tag{11}$$

As an individual-oriented intervention decreases the variance in the beliefs about θ, it decreases the weight players place on s_t, thereby making the decision rule more sensitive to v_t and establishing the comparative static in the lemma. This increased sensitivity implies that

giving information about θ increases the endogenous reputational penalty from adapting to one's private information and thus never breaks a pooling equilibrium.

In contrast, a peer-oriented intervention may break a pooling equilibrium. One can observe that the left-hand side of the condition in the lemma is increasing in $\sigma^2_{\theta,t}$ and the right-hand side is increasing in $\sigma^2_{\mu,t}$ (because in any signaling equilibrium, the slope of the beliefs about a player's type with respect to their decision, α, is positive). Therefore, a peer-oriented intervention will reduce the reputational penalty from adaptation whenever the uncertainty with respect to θ is comparatively larger than the uncertainty with respect to μ. To see why, note that providing information about μ has two effects. First, each player wants to be perceived as μ. Consequently, when $\sigma^2_{\mu,t}$ is high, 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 $\sigma^2_{\mu,t}$ is high, the players have an added incentive to adapt, resulting in revelation.

The countervailing effect is that when $\sigma^2_{\mu,t}$ is high, 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 value of $\sigma^2_{\theta,t}$ has no effect on the first force but does affect the latter. To see why $\sigma^2_{\theta,t}$ impacts the latter force, note that when $\sigma^2_{\theta,t}$ is low, 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 a decrease in $\sigma^2_{\mu,t}$. As a result, when $\sigma^2_{\theta,t}$ is low, decreasing $\sigma^2_{\mu,t}$ 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.

Further, 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 build intuition for the forces behind the change, I consider the following situation: suppose the prior beliefs about μ, θ 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:

$$a^*(a_1) = \mathbb{E}(\theta|a_1) + \mathbb{E}(\mu|a_1). \tag{12}$$

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:

$$a_1 = \lambda_s s_1 + \lambda_v v_1, \tag{13}$$

for some λ_s, λ_v. I now consider the following intervention where the public history is adapted to be either $h^2(\theta) = \{a_1, \gamma_1, \theta\}$ or $h^2(\mu) = \{a_1, \gamma_1, \mu\}$ and analyze how the pooling decision rule changes given this information. Suppose that the players utilize the decision-rule in Equation (13), resulting in a pooling decision rule denoted by $a^*(a_1)$ as in Equation (12) for all subsequent periods.

Proposition 4 (**Effect of Interventions on Pooling Decision**)
Denote by $\Delta(\theta)$ (respectively, $\Delta(\mu)$) the difference between the new decision the players pool on compared to $a^*(a_1)$. Then,

$$\Delta(\mu) = \Big(1 - \frac{\sigma_\theta^2 \, \lambda_s \lambda_v}{\lambda_v^2 + \lambda_s^2(\sigma_\theta^2 + 1)}\Big)\mu + c_\mu a_1 \tag{14}$$

$$\Delta(\theta) = \Big(1 - \frac{\sigma_\mu^2 \, \lambda_s \lambda_v}{\lambda_s^2 + \lambda_v^2(\sigma_\mu^2 + 1)}\Big)\theta + c_\theta a_1 \tag{15}$$

for some constants c_θ, c_μ.

To understand the expressions above, let us now consider the effect of revealing θ, where a symmetric analysis occurs for μ. Trivially, $\mathbb{E}(\theta|h^2(\theta)) = \theta$. Further, the players re-evaluate their perception of μ as a function of both θ (the second term in the parentheses of Equation (15)) and a_1. The object of interest is how much the players' decision changes with respect to θ. One can see that given information that θ is positive (respectively, negative) the players update their belief 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 σ_μ^2 is large (i.e., the players are uncertain about their peers' true preferences). Such a counter-update provides a rationale for why individual-oriented interventions may have small (if not negative) effects on behavior.

6 Empirical Predictions

This section relates the theoretical findings in this paper to key empirical findings. Table 1 below discusses how the various settings described in the introduction explicitly map into the model. The first subsection discusses how the learning predictions in Section 4 relate to empirical findings on social learning. The second subsection discusses how the intervention results connect to the empirical literature.

Table 1: Empirical Literature in Support of This Modeling

Context	**Interpretation within My Model**	**Results Consistent With My Model**	**Social Learning Literature Prediction**
Alcohol Consumption	θ = relative safety of alcohol consumption, μ = peers' attitudes towards alcohol, $\kappa > 0$ as decisions are public	Schroeder and Prentice (1998) show information about μ is more effective than information about θ.	Don't model μ; therefore information about θ is, tautologically, optimal.
Women Working Outside the Home (WWOH)	θ = economic benefits of WWOH, μ = peers' attitudes towards WWOH, $\kappa > 0$ as decisions are public	Bursztyn et al. (2020a) document misperceptions of μ and correcting misperceptions of μ increases WWOH	Don't model μ, therefore information about θ is, tautologically, optimal.
Political Speech	θ = optimal policy, μ = peers' political positions, $\kappa \simeq 0$ when said in private, $\kappa > 0$ when said in public	Braghieri (2021) shows speech is less informative in public than private. Bursztyn and Yang (2022) summarize misperceptions of peers' political positions.	Speech is "responsive" to private information, as one can state "I am 94.32% confident in this policy," so there should be no misperceptions (Ali, 2018).
Political Actions	θ = optimal policy, μ = peers' political positions, $\kappa \simeq 0$ when said in private, $\kappa > 0$ when said in public	Bursztyn et al. (2020b) show information about Trump's popularity increased donations to xenophobic charities, but only when donations were public.	No difference between public and private. Only possible impact of information would be updated beliefs about whether xenophobia is an optimal policy.
Medical Behaviors	θ = healthy policy, μ = peers' attitudes towards these behaviors, $\kappa > 0$ as choices are observed	Ferreira et al. (2024) show Female Genital Cutting influenced by incorrect perceptions of μ.	Only way to influence medical behavior is by beliefs about θ.
Friendships	θ = optimal empathy level, μ = peers' true empathy levels, $\kappa > 0$ as decisions are public	Pei et al. (2025) show students misperceive their peers' empathy levels and correcting these misperceptions leads to expanded social networks and social risk-taking.	Only way to influence the "loneliness epidemic" on college campuses is by beliefs about θ.

6.1 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: The standard social learning literature predicts that with continuous actions, and under mild regularity conditions, behavior should converge to efficiency in the long run (Lee, 1993; Ali, 2018; Kartik et al., 2024). Yet, as Prentice and Miller (1993) show in the context of alcohol consumption, arguably an environment with continuous decisions, systematic misperceptions can persist in equilibrium, leading to inefficiency. This outcome is consistent with my model, as this environment also exhibits large conformity motives, and thus the model predicts misperceptions.

Another environment consistent with my findings is speech. Consider individuals on a social media platform who provide their opinion on a given topic. Individual t incurs a lying cost if his stated opinion, a_t, deviates too far from his true opinion on the topic, $\theta + v_t$. Here, θ corresponds to the objective component of an issue (e.g., its factual or economic consequences), and v_t captures an idiosyncratic, purely subjective taste. This environment is continuous, as individuals can make statements such as "I am x% in support of this policy" for any x. While standard models would predict efficient aggregation of information, my model predicts that such aggregation occurs only for topics with low conformity concerns, where the incentive to shade one's report toward the perceived consensus is small. For topics with high conformity concerns (e.g., politically sensitive issues), publicly stated opinions will fail to aggregate individuals' private information. The analysis in Braghieri (2021) provides broad support for this prediction and shows that public speech is informative about private opinions only for non-politically sensitive topics.[14] My results also suggest that the analysis in Braghieri (2021) is robust to his modeling of communication using a discrete response scale (in his experiment the value of x can take only one of seven points). In my framework, unlike the remainder of the social learning literature, the qualitative results do not depend on this discrete elicitation: allowing individuals to report any x does not restore aggregation when conformity concerns are high. Finally, as I show formally in Section 7, conformity concerns will also prevent individuals from aggregating information about the average opinion in the group even absent uncertainty about θ (e.g., everyone knows the economic implications of a given policy).

[14] Braghieri (2021) includes a model which has two key differences from mine: (i) actions are discrete and (ii) the correct perception is commonly known, whereas I allow for uncertainty about the optimal perception, allowing for pluralistic ignorance, discussed below.

Pluralistic Ignorance: My model predicts that "pluralistic ignorance" can arise in equilibrium. Miller and Prentice (1994) define 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.

A common empirical measure of pluralistic ignorance is the gap between the group's true average preference type and the average of individuals' perceptions of that preference type. Because this is measured for a fixed group at a fixed time, a natural model analogue at history h^t is $|\mu - \mathbb{E}(\mu|\mu(h^t), v_t)|$.[15] If learning occurs, then tautologically every player correctly estimates μ. In contrast, if learning fails (i.e., $\mu(h^t) \not\to \mu$), the uncertainty about μ implies that each player's estimate of μ places weight both on the expectation of μ given $\mu(h^t)$, $\bar{\mu}(h^t)$, and on his own preference type v_t, which is informative about μ. Moreover, when $\mu(h^t)$ is imprecise, $\bar{\mu}(h^t)$ is likely to be far away from μ, causing correlated (i.e., "systematic") misperceptions across individuals. Further, the greater the uncertainty in the public beliefs, $\mu(h^t)$, the further $\bar{\mu}(h^t)$ is from μ, resulting in greater pluralistic ignorance on average. Finally, 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" (Loury, 1994), resulting in insufficient information for others to gauge the views of the public.

6.2 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 3 suggests that peer-oriented interventions may be

[15] That $\mu(h^t)$ is an unbiased estimator for μ does not rule out large realized errors for a given group at a given time, which is the empirical object social scientists measure. One may object that, in practice, beliefs about peer preferences can be biased in the same direction across many groups (e.g., students systematically overestimate peers' enthusiasm for alcohol). This is consistent with the model once one recognizes that different schools may be exposed to similar media environments (celebrities on social media or television drinking), so beliefs across schools should not be viewed as independent draws.

preferred because they can break a pooling equilibrium. Further, Proposition 4 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), Ferreira et al. (2024), and Pei et al. (2025) 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.

Finally, this analysis has implications for interventions more broadly. Because the reputational penalty from adapting to one's private information is higher when actions are more informative about an individual's preference type, mechanisms that make individual choices less informative can result in social learning by lowering the reputational cost of deviating from the norm. For instance, in the context of alcohol consumption, the increasing prevalence of non-alcoholic beers allow individuals to consume less alcohol without reputational concerns.[16] Additionally, when the conformity concerns are large, instituting private and anonymous decisions (which, tautologically, prevent social learning) may be preferable, as private decision-making allows individuals to adapt their actions to their signals and types, instead of pooling when decisions are public. In contrast, social learning models absent conformity concerns tend to feature the improvement principle: as individual $t+1$ can always mimic individual t's action, his expected utility must be weakly higher. Therefore, instituting private decision-making does not improve the expected utility of the players.

7 Extensions and Comparative Statics

In this section, I derive additional predictions from Section 4 by restricting attention to the special cases where at least one of θ, μ is common knowledge. Equivalently, one can view these analyses as the case where the prior variance is equal to zero, and by continuity, the qualitative results would continue to hold with small amounts of uncertainty. The first subsection derives additional results for the case in which both θ, μ are common knowledge. Subsection two considers the case where θ is common knowledge but μ is not, and Subsection

[16] Similarly, Boudreau et al. (2023) show that adding noise to harassment reporting can make reports more informative about a player's private information by breaking a pooling equilibrium of no reporting. One can see the same comparative static in Lemma 3: when there is more variance in a_t coming from noise (in my model, differential beliefs about θ), it is easier to sustain an equilibrium with revelation.

three considers the parallel case. In each subsection, I will also provide results beyond the Linear-Gaussian setting analyzed in the main analysis.

7.1 Common Knowledge Benchmark Revisited

This subsection analyzes Perfect Bayesian Equilibria when θ and μ are common knowledge, as in Section 3. Similarly to Section 3, it is without loss of generality to assume θ, μ are equal to zero. Further, without uncertainty, it is without loss of generality to consider the decision rule of a player with preference type v.

Notably, the linear equilibria derived in Section 3 continue to exist for any distribution of v that admits a full-support density, 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 (**Central Pooling 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:*

$$a'(v) = \frac{\kappa v}{v - a(v)} > 0. \tag{16}$$

Proof of Lemma 4. The statement follows from Theorem 3 in Bernheim (1994). □

The intuition behind this 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 (16). Given this differential equation, one can prove the following proposition.

Proposition 5 (**Fully Pooling Equilibria**)
When $\kappa > \kappa^{c.k.}$, *no equilibrium with revelation satisfies D1. If* $\kappa \leq \kappa^{c.k.}$, *the revealing linear equilibria satisfy D1.*

The above result was not present in Bernheim (1994), but mirrors the intuition of the

case with linear equilibria.[17] As the conformity concerns increase, any revealing decision rule flattens as all individuals choose less extreme decisions. These flatter decision rules increase the incentive for extreme types to deviate and choose a conforming decision as their equilibrium decision already incurs a large adaptation loss. Therefore, there is a cutoff where the decision rule cannot become any flatter without generating too large of a deviation temptation for such players, precluding an equilibrium with revelation. Finally, the revealing linear equilibria trivially satisfy D1 as no actions are off path.[18]

7.2 Learning Peers' Preferences

Here, θ is commonly known (i.e., its prior variance is zero), and I analyze whether players learn μ. Denote by $g(v_t|h^t) := \mathbb{E}(\mu|v_t, h^t)$. I make two assumptions: (i) the distribution of v_t admits a full support density and (ii) $g'(v_t|h^t) \in (0,1)$ and is continuous. These conditions hold for all h^t whenever $v_t = \mu + \nu_t$ where ν_t is i.i.d. across time and mean zero and both ν_t, μ admit continuously differentiable and log-concave densities with full support. Notably, the Gaussian framework in Section 2 satisfies these assumptions. Given these assumptions, one can show the following result.

Proposition 6 (**General Distributions with θ known**)
Fix a history h^t and denote by $\underline{g}(h^t) := \inf_x g'(x|h^t) \leq \sup_x g'(x|h^t) := \bar{g}(h^t)$, where $g(x|h^t) = \mathbb{E}(\mu|x, h^t)$. There exists a fully revealing equilibrium that satisfies D1 if the conformity concerns, κ, satisfy

$$\kappa \leq \frac{\kappa^{c.k.}}{1 - \underline{g}(h^t)}. \tag{17}$$

Further, no equilibrium with revelation satisfies D1 if

$$\kappa > \frac{\kappa^{c.k.}}{1 - \bar{g}(h^t)}. \tag{18}$$

Therefore, there exists an equilibrium that satisfies D1 with social learning about preferences if and only if $\kappa \leq \kappa^{c.k.}$.

[17] The lone difference to Bernheim (1994) is that I consider an unbounded support for v, whereas Bernheim (1994) considers a bounded support. As can be seen in the proof, the qualitative results do not rely on this distinction.

[18] The pooling linear equilibria are trivially central pooling (by setting $-\underline{v} = \bar{v} = \infty$). Furthermore, these equilibria satisfy the D1 Criterion under a compactification of the type space to $\text{Supp}(v) = [-\infty, \infty]$. Under this refinement, off-path beliefs are supported on the boundary types ($\pm\infty$), which represent the supremum of incentives to deviate. Finally, all results presented in this paper remain robust to this extension of D1.

The intuition behind this proposition can be seen from Lemma 3. As there is no uncertainty about θ, greater uncertainty about μ makes it easier to sustain a revealing equilibrium. First, each player wants to be perceived as μ. Consequently, when uncertainty about μ is high, 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. Further, one can view $\bar{g}(h^t), \underline{g}(h^t)$ as some measures of uncertainty. When μ is very uncertain, a player's posterior beliefs about μ are more sensitive to their own type, implying $g(\cdot|h^t)$ has a higher slope. Therefore, when $\bar{g}(h^t), \underline{g}(h^t)$ are higher it is easier to sustain an equilibrium with revelation, as reflected in the proposition. Further, if $\kappa \leq \kappa^{\text{c.k.}}$ then the sufficient condition for the existence of a fully revealing equilibrium holds for every history. In this equilibrium, $\mu(h^t) \to_p \mu$ by the Law of Large Numbers. In contrast, if $\kappa > \kappa^{\text{c.k.}}$, then $\bar{g}(h^t) \not\to_p 0$ in any equilibrium, precluding social learning about μ. These results imply that even beyond the Linear-Gaussian environment, the magnitude of conformity concerns continues to predict learning outcomes.

This proposition gives a separate necessary and sufficient condition for the existence of an equilibrium with revelation. In the linear Gaussian analysis, $g(\cdot|h^t)$ is linear for all t, implying that $\bar{g}(h^t) = \underline{g}(h^t)$ and that the bounds in the proposition are tight. Further, in such an analysis $\bar{g}(h^t)$ is a deterministic function of t and decreases if and only if the previous period involved revelation. This determinism implies that in the signaling equilibrium, the asymptotic variance in the beliefs about μ is increasing in κ, and therefore the asymptotic utility and adaptation loss are decreasing in κ. The following corollary formalizes this intuition when analyzing the core model in Section 2, but restricting attention to $\sigma_\theta^2 = 0$.

Corollary 1 (**Signaling Equilibrium with θ known (i.e., $\sigma_\theta = 0$)**)
Social learning about fundamentals occurs if and only if $\kappa \leq \kappa^{c.k.}$. Further, when $\kappa > \kappa^{c.k.}$ the asymptotic variance of the beliefs, $\lim \sigma_{\mu,t}^2$, is positive and increasing in κ. Finally, the asymptotic adaptation loss and asymptotic utility of the players is decreasing in κ.

7.3 Learning Economic Fundamentals

This subsection assumes that the prior variance of μ is equal to zero and analyzes whether the players learn θ. As the learning is univariate, one can normalize $\gamma_t = 1$ without loss of insights (see Subsection 2.1). In this case, a player's type is, without loss of generality, equal to $\tilde{v}_t := v_t + \mathbb{E}(\theta|s_t, h^t)$, which I assume admits a continuous density with full support. Define $m(\tilde{v}_t|h^t) := \mathbb{E}(v_t^2|\tilde{v}_t, h^t)$. This measures the conformity loss from being revealed to have type $\tilde{v}_t$ at history h^t, and depends only on the joint distribution of $(v_t, \tilde{v}_t)$ given h^t, not

on equilibrium play in that period. Assume that for every history h^t, the function $m(\cdot \mid h^t)$ is twice differentiable and satisfies $0 \leq m''(x \mid h^t) \leq 2$ for all $x \in \mathbb{R}$.[19] Intuitively, when a player is revealed to have one-dimensional type $\tilde{v}_t$ close to $\mathbb{E}(\theta|h^t)$, this is most consistent with a small v_t, so the expected conformity loss $m(\cdot|h^t)$ is minimized at $\mathbb{E}(\theta|h^t)$. By contrast, observing $\tilde{v}_t$ far away from $\mathbb{E}(\theta|h^t)$ is typically explained by a more extreme realization of v_t, hence the convexity.

Proposition 7 (General Distributions with μ known)
Fix a history h^t and denote by $\underline{m}(h^t) := \inf_x m''(\tilde{v}_t|h^t)/2 \leq \sup_x m''(\tilde{v}_t|h^t)/2 := \bar{m}(h^t)$, where $\tilde{v}_t = v_t + \mathbb{E}(\theta|h^t, s_t)$ and $m(\tilde{v}_t|h^t) = \mathbb{E}(v_t^2|\tilde{v}_t, h^t)$. There exists a fully revealing equilibrium that satisfies D1 if the conformity concerns, κ, satisfy

$$\kappa \leq \frac{\kappa^{c.k.}}{\bar{m}(h^t)}. \tag{19}$$

Further, no equilibrium with revelation satisfies D1 if

$$\kappa > \frac{\kappa^{c.k.}}{\underline{m}(h^t)}. \tag{20}$$

Therefore, there exists an equilibrium that satisfies D1 with social learning about fundamentals if and only if $\kappa \leq \kappa^{c.k.}$.

The intuition behind this proposition can be seen from Lemma 3. When there is more uncertainty about θ, the sensitivity decreases as players lose plausible deniability for choosing different decisions, making it harder to sustain an equilibrium with revelation. Further, one can view $\bar{m}(h^t), \underline{m}(h^t)$ as some measures of uncertainty. When θ is very uncertain, $\tilde{v}_t$ includes a lot of noise from the player's posterior estimate of θ. Therefore, the conformity loss, as measured by $m(\cdot|h^t)$ is less sensitive to the decision, resulting in a lower curvature of the reputational loss. Therefore, when $\bar{m}(h^t), \underline{m}(h^t)$ are lower it is easier to sustain an equilibrium with revelation, as reflected in the proposition. Further, if $\kappa \leq \kappa^{\text{c.k.}}$ then the sufficient condition for the existence of a fully revealing equilibrium holds for every history. In this equilibrium, $\theta(h^t) \to_p \theta$ since infinitely many informative signals are generated. In contrast, if $\kappa > \kappa^{\text{c.k.}}$, there must exist a finite time for which from then on the players are pooling. These results imply that even beyond the Linear-Gaussian environment, the magnitude of conformity concerns continues to predict learning outcomes.

[19] That $\tilde{v}_t$ admits a differentiable density holds whenever the distributions of θ, v_t, ϵ_t admit full support and differentiable densities. Further the convexity requirement holds in the limit of no uncertainty: $m(v_t) = v_t^2 \implies m''(v_t) = 2$. By continuity, derivative will remain positive everywhere and the noise from θ will dampen the reputational inference, causing a slight decrease in the curvature.

This proposition gives a separate necessary and sufficient condition for the existence of an equilibrium with revelation. In the linear Gaussian analysis, $m(\cdot|h^t)$ is linear for all t implying that $\bar{m}(h^t) = \underline{m}(h^t)$ and that these bounds are tight. Further, in such an analysis $\bar{m}(h^t)$ is a deterministic function of t and increases if and only if the previous period involved revelation. This determinism implies that in the signaling equilibrium, the asymptotic variance in the beliefs about θ is increasing in κ, implying the asymptotic utility and adaptation loss are decreasing in κ. The following corollary formalizes this intuition when analyzing the core model as in Section 2, but restricting attention to $\sigma_\mu^2 = 0$.

Corollary 2 (**Signaling Equilibrium with μ known (i.e., $\sigma_\mu = 0$)**)
Social learning about fundamentals occurs if and only if $\kappa \leq \kappa^{c.k.}$. Further, when $\kappa > \kappa^{c.k.}$ the asymptotic variance of the beliefs $\lim \sigma_{\theta,t}^2 > 0$ is increasing in κ. Finally, the asymptotic adaptation loss and asymptotic utility of the players is decreasing in κ.

8 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 frame-

work 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.

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

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

Ball, Ian, "Scoring strategic agents," *arXiv preprint arXiv:1909.01888*, 2019.

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.

Boudreau, Laura E, Sylvain Chassang, Ada Gonzalez-Torres, Rachel Heath, and National Bureau of Economic Research, "Monitoring harassment in organizations," Technical Report, National Bureau of Economic Research Cambridge, MA 2023.

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 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.

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

Echenique, Federico, "Comparative statics by adaptive dynamics and the correspondence principle," *Econometrica*, 2002, *70* (2), 833–844.

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.

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

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.

Holmström, Bengt, "Managerial incentive problems: A dynamic perspective," *The review of Economic studies*, 1999, *66* (1), 169–182.

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.

Pei, Rui, Samantha J Grayson, Ruth E Appel, Serena Soh, Sydney B Garcia, Annabel Bouwer, Emily Huang, Matthew O Jackson, Gabriella M Harari, and Jamil Zaki, "Bridging the empathy perception gap fosters social connection," *Nature Human Behaviour*, 2025, pp. 1–14.

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.

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.

9 Appendix A

Proof of Proposition 1. By Lemma 1 a sufficient condition for the existence of a revealing equilibrium absent uncertainty is the solution to $1 + \kappa\alpha^2 = \alpha$. This equation has a solution if and only if $\kappa \leq 1/4$, proving the existence of $\kappa^{\text{c.k.}}$. When $\kappa > \kappa^{\text{c.k.}}$, the players must be pooling and $a = 0$ is the pooling decision which satisfies the Pareto refinement. When $\kappa \leq \kappa^{\text{c.k.}}$, the revealing equilibrium has $a(v) = v/\alpha$, where α solves $1 + \kappa\alpha^2 = \alpha$. As $\alpha \geq 1$, the equilibrium with the smallest positive value of α, results in the decision rule which minimizes the adaptation loss. As the conformity loss is fixed by the Law of Iterated Expectation, this decision rule is selected by the Pareto refinement. Algebraic simplifications then derive the results in the proposition. □

Proof of Lemma 1. Denote by Σ_t the variance matrix of the beliefs in period t

$$\Sigma_t = \begin{pmatrix} \sigma^2_{\theta,t} & \rho_t\sigma_{\theta,t}\sigma_{\mu,t} \\ \rho_t\sigma_{\theta,t}\sigma_{\mu,t} & \sigma^2_{\mu,t} \end{pmatrix} \Bigg(\tag{21}$$

Given the Gaussian uncertainty, the posterior means of (θ_t, μ_t) given (s_t, v_t) are linear:

$$\begin{pmatrix} \mathbb{E}(\theta \mid s_t, v_t, \theta(h^t), \mu(h^t)) \\ \mathbb{E}(\mu \mid s_t, v_t, \theta(h^t), \mu(h^t)) \end{pmatrix} = K_t \begin{pmatrix} s_t \\ v_t \end{pmatrix} + (\Sigma_t + I)^{-1} \begin{pmatrix} \bar{\theta}_t \\ \bar{\mu}_t \end{pmatrix}, \text{with } K_t := \Sigma_t(\Sigma_t + I)^{-1}, \tag{22}$$

and I is the two-by-two identity matrix. One can simplify to show that

$$K_t = \begin{pmatrix} \dfrac{\sigma^2_{\theta,t}(1+\sigma^2_{\mu,t}) - \rho_t^2\sigma^2_{\theta,t}\sigma^2_{\mu,t}}{(1+\sigma^2_{\theta,t})(1+\sigma^2_{\mu,t}) - \rho_t^2\sigma^2_{\theta,t}\sigma^2_{\mu,t}} & \dfrac{\rho_t\sigma_{\theta,t}\sigma_{\mu,t}}{(1+\sigma^2_{\theta,t})(1+\sigma^2_{\mu,t}) - \rho_t^2\sigma^2_{\theta,t}\sigma^2_{\mu,t}} \\ \dfrac{\rho_t\sigma_{\theta,t}\sigma_{\mu,t}}{(1+\sigma^2_{\theta,t})(1+\sigma^2_{\mu,t}) - \rho_t^2\sigma^2_{\theta,t}\sigma^2_{\mu,t}} & \dfrac{\sigma^2_{\mu,t}(1+\sigma^2_{\theta,t}) - \rho_t^2\sigma^2_{\theta,t}\sigma^2_{\mu,t}}{(1+\sigma^2_{\theta,t})(1+\sigma^2_{\mu,t}) - \rho_t^2\sigma^2_{\theta,t}\sigma^2_{\mu,t}} \end{pmatrix} \Bigg(\tag{23}$$

The proofs below do not require a closed form for K_t; I will use only the following remarks:

Remark 1 $(K_t)_{1,1} \geq 0$*, is continuous in* $\sigma^2_{\theta,t}$*, and equals zero if and only if* $\sigma^2_{\theta,t} = 0$*.*

This states that the posterior expectation of θ is linear in s_t with a positive slope, and that this slope is increasing in the uncertainty about θ. Remark 2 shows a similar result for μ.

Remark 2 $(K_t)_{2,2} \geq 0$*, is continuous in* $\sigma^2_{\mu,t}$*, and equals zero if and only if* $\sigma^2_{\mu,t} = 0$*.*

Finally, if beliefs about θ, μ are correlated, then s_t predicts μ and v_t predicts θ. The remarks below bound the strength of this predictability.

Remark 3 $(K_t)_{1,2} = (K_t)_{2,1}$, $|(K_t)_{1,2}| \leq 1/2$. *Further,* $(K_t)_{1,2}$ *is continuous in* $\sigma^2_{\mu,t}, \sigma^2_{\theta,t}$ *and equals zero if either* $\sigma^2_{\mu,t}, \sigma^2_{\theta,t}$ *equals zero. Finally, the sign of* $(K_t)_{1,2}$ *equals the sign of* ρ_t.

Fix a conjectured belief $\hat{v}(t) = \alpha_t a_t + \beta_t$. The first-order condition given this conjecture is Equation (6). By Equation (22), for some exogenous constants, $c_{1,t}, \ldots, c_{6,t}$,

$$a_t(1 + \kappa\alpha_t^2) = -\kappa\alpha_t\beta_t + c_{1,t} + c_{2,t}\alpha_t\kappa + s_t(c_{3,t} + \kappa\alpha_t c_{4,t}) + v_t(1 + c_{5,t} + \kappa\alpha_t c_{6,t}) := y_t \quad (24)$$

Here, $c_{1,t}, c_{2,t}$ correspond to the coefficients multiplying $\bar{\theta}_t, \bar{\mu}_t$ in Equation (22). Further, $c_{3,t} = \gamma_t(K_t)_{1,1}, c_{4,t} = (K_t)_{1,2}, c_{5,t} = \gamma_t(K_t)_{2,1}, c_{6,t} = (K_t)_{2,2}$. Therefore,

$$\begin{aligned}
\mathbb{E}(v_t|y_t, \theta(h^t), \mu(h^t)) &= \bar{\mu}(t) + \frac{\mathrm{Cov}(v_t, y_t|\theta(h^t), \mu(h^t))}{\mathrm{Var}(y_t|\theta(h^t), \mu(h^t)}\Big(y_t - \mathbb{E}\big(y_t|\theta(h^t), \mu(h^t)\big)\Big) \\
&:= \bar{\mu}(t) + \tilde{s}(\theta(h^t), \mu(h^t), \alpha, \kappa)\Big(y_t - \mathbb{E}\big(y_t|\theta(h^t), \mu(h^t)\big)\Big) \\
&= \bar{\mu}(t) + \tilde{s}(\theta(h^t), \mu(h^t), \alpha, \kappa)(1 + \kappa\alpha_t^2)\Big(a_t - \mathbb{E}\big(a_t|\theta(h^t), \mu(h^t)\big)\Big), \quad (25)
\end{aligned}$$

where one can observe that $\tilde{s}(\cdot)$ does not depend on β_t, $\bar{\theta}(t), \bar{\mu}(t)$, or the realizations of θ, μ.

Therefore, a necessary condition for the existence of a revealing linear equilibrium is that $\alpha_t a_t + \beta_t$ is equal to Equation (25) for all on-path a_t. Further, this equation is sufficient because, as argued in the text, $\hat{v}(a_t)$ is a sufficient statistic for $\phi(b|a_t, h^t)$ in determining the optimal decision of the players. Therefore, fixing any values of α_t, β_t which satisfy this equation, then $\mathrm{Var}(v_t|a_t, h^t)$ is uniquely pinned down by (24), and given that $\phi(b|a_t, h^t)$ is Gaussian, $\mathrm{Var}(v_t|a_t, h^t)$ and $\mathbb{E}(v_t|a_t, h^t)$ uniquely determine this distribution. Finally, in any linear revealing equilibrium, all a_t are on path implying the following is necessary and sufficient for a revealing linear equilibrium to exist:

$$\alpha_t = (1 + \kappa\alpha_t^2)\tilde{s}(\theta(h^t), \mu(h^t), \alpha, \kappa) \quad (26)$$

$$\beta_t = \bar{\mu}(t) - \tilde{s}(\theta(h^t), \mu(h^t), \alpha, \kappa)(1 + \kappa\alpha_t^2)\mathbb{E}\big(a_t|\theta(h^t), \mu(h^t)\big). \quad (27)$$

Note that (26) does not contain β_t. I now argue that for any solution α_t to (26) there is a unique solution to (27) and hence β_t. Given a solution to (26), (27) simplifies to

$$\begin{aligned}
&\beta_t = \bar{\mu}(t) - \frac{\alpha_t}{1 + \kappa\alpha_t^2}\Big(-\kappa\alpha_t\beta_t + c_{1,t} + c_{2,t}\alpha_t\kappa + \bar{\theta}(t)(c_{3,t} + \kappa\alpha_t c_{4,t}) + \bar{\mu}(t)(1 + c_{5,t} + \kappa\alpha_t c_{6,t})\Big) \\
&\iff \beta_t\frac{1}{1 + \kappa\alpha_t^2} = \bar{\mu}(t) - \alpha_t\Big(c_{1,t} + c_{2,t}\alpha_t\kappa + \bar{\theta}(t)(c_{3,t} + \kappa\alpha_t c_{4,t}) + \bar{\mu}(t)(1 + c_{5,t} + \kappa\alpha_t c_{6,t})\Big).
\end{aligned}$$

However, as the coefficient on β_t is non-zero, any solution to (26) results in a unique solution for (27). Further, as the conformity loss is fixed in any equilibrium by the Law of Iterated

Expectations, the equilibrium which maximizes expected utility is the one which minimizes the adaptation loss. Therefore, if an equilibrium with revelation exists, then the signaling equilibrium will be revealing. □

I now present three lemmas and one definition not stated in the main text.

Definition 5 *Denote by* $r(\alpha|h^t) = (1+\kappa\alpha^2)\text{Cov}(y_t, v_t|\theta(h^t), \mu(h^t)) - \alpha\text{Var}(y_t|\theta(h^t), \mu(h^t))$, *where* y_t *was defined in Equation* (24).

Given the definition of the sensitivity, $r(\alpha|h^t)$ has a root whenever Equation (8) has a root.

Lemma A1 *For any history,* h^t, $\lim_{\alpha\to\infty} r(\alpha|h^t) = +\infty$.

Proof of Lemma A1. Given the definition of y_t, $r(\alpha|h^t)$ is a cubic polynomial in α with leading coefficient equal to

$$\kappa^2\text{Cov}\left(\mathbb{E}\left(\mu|v_t, s_t\right), v_t\right) - \kappa^2\text{Var}\left(\mathbb{E}\left(\mu|v_t, s_t\right)\right). \tag{28}$$

As we are interested in the sign, we can drop the κ^2. Further, by definition $v_t = \nu_t + \mu$, thus the sign of the leading coefficient is equal to the sign of

$$\text{Cov}\left(\mathbb{E}\left(\mu|v_t, s_t\right), \nu_t\right) + \text{Cov}\left(\mathbb{E}\left(\mu|v_t, s_t\right), \mu\right) - \text{Var}\left(\mathbb{E}\left(\mu|v_t, s_t\right)\right) = \text{Cov}\left(\mathbb{E}\left(\mu|v_t, s_t\right), \nu_t\right),$$

which follows as μ equals $\mathbb{E}(\mu|v_t, s_t)$ plus noise which is independent of $\mathbb{E}(\mu|v_t, s_t)$, implying the latter two terms cancel. Therefore, it suffices to analyze the sign of $\text{Cov}\left(\mathbb{E}\left(\mu|v_t, s_t\right), \nu_t\right)$. Finally, $\mathbb{E}\left(\mu|v_t, s_t\right) = c_{6,t}v_t + c_{4,t}s_t$, and ν_t is independent of s_t, as defined in Equation (24). Therefore, $\text{Cov}\left(\mathbb{E}\left(\mu|v_t, s_t\right), \nu_t\right) = c_{6,t} \cdot \text{Cov}(\nu_t, v_t)$. Finally, observe that this covariance is positive and $c_{6,t} = (K_t)_{2,2} > 0$ by Remark 2. □

Lemma A2 *The signaling equilibrium is revealing if and only if*

$$\begin{aligned} 0 &\geq \inf_{\alpha\geq 0}(1+\kappa\alpha^2)\tilde{s}(\theta(h^t), \mu(h^t), \alpha, \kappa) - \alpha \iff \\ 0 &\geq \inf_{\alpha\geq 0}(1+\kappa\alpha^2)\text{Cov}(y_t, v_t|\theta(h^t), \mu(h^t)) - \alpha\text{Var}(y_t|\theta(h^t), \mu(h^t)). \end{aligned} \tag{29}$$

Proof of Lemma A2. The first and second expressions in Equation (29) are equivalent by the definition of the sensitivity and from noting that $\text{Var}(y_t|\theta(h^t), \mu(h^t)) > 0$. Further, by Lemma 1, a root to the expression in Equation (29) is sufficient for the signaling equilibrium to be revealing. Finally, as these expressions are continuous and Lemma A1 implies they take positive values as $\alpha \to \infty$, the existence of a root is equivalent to the existence of a negative value. □

With an abuse of notation, I will drop the dependence of the covariance and variance operators in Equation (29) on $\theta(h^t), \mu(h^t)$.

Lemma A3 *Fix any revealing subsequence $\{t_i\}$ then α_{t_i} is bounded.*

Proof of Lemma A3. Proposition A4 in the Online Appendix implies $\Sigma_{t_i} \to \bar{\Sigma}$, so the coefficients of $r_{t_i}(\alpha) = a_{3,i}\alpha^3 + a_{2,i}\alpha^2 + a_{1,i}\alpha + a_{0,i}$ are bounded. By Lemma A1, $a_{3,i} = \kappa^2 (K_{t_i})_{2,2} \geq 0$. If $\liminf_i (K_{t_i})_{2,2} > 0$, then $a_{3,i}$ is bounded away from 0 and a Cauchy root bound implies α_{t_i} is uniformly bounded.

Otherwise $\sigma_{\mu,t_i} \to 0$, hence $c_{4,t_i}, c_{5,t_i}, c_{6,t_i} \to 0, \text{Cov}(s_{t_i}, v_{t_i}) \to 0$ and $\text{Var}(v_{t_i}) \to 1$. The quadratic coefficient is

$$a_{2,i} = \kappa\big(c_{3,t_i}\text{Cov}(s_{t_i}, v_{t_i}) + (1 + c_{5,t_i})\text{Var}(v_{t_i})\big) \Bigg(\\ - 2\kappa\big(c_{3,t_i}c_{4,t_i}\text{Var}(s_{t_i}) + (1 + c_{5,t_i})c_{6,t_i}\text{Var}\Big(v_{t_i}) + (c_{3,t_i}c_{6,t_i} + c_{4,t_i}(1 + c_{5,t_i}))\text{Cov}(s_{t_i}, v_{t_i})\big), \Bigg($$

so $a_{2,i} \to \kappa > 0$. As the cubic coefficient is weakly positive, the quadratic coefficient is strictly positive, and all other coefficients are bounded, the roots must also be bounded. □

Proof of Lemma 2. I first derive a sufficient condition for the equilibrium to be pooling in the limit, thus precluding social learning about either fundamentals or preferences. Next, I show that if, by contradiction, either social learning about fundamentals or preferences occurs, that this condition holds, deriving the contradiction.

Claim 1: If $|c_{3,t}|, |c_{4,t}|, |c_{5,t}|, |c_{6,t}| \to 0$, where such constants are defined in Equation (24), then there exists a T, such that for $t \geq T$, the signaling equilibrium is pooling.

Proof of Claim 1: Suppose not, then there exists a subsequence t_i for which the signaling equilibrium is revealing in period t_i. By Lemma A2, $r(\alpha|h^{t_i})$ has a positive root α_{t_i} for all t_i. Further, given that $|c_{3,t_i}|, |c_{4,t_i}|, |c_{5,t_i}|, |c_{6,t_i}| \to 0$, the cubic polynomial, $r(\alpha|h^{t_i})$, pointwise converges to $1 + \kappa\alpha^2 - \alpha$ as $t_i \to \infty$. Further, as stated in Lemma A1 the leading coefficient of this cubic is positive for all t_i. Therefore, for large enough t_i, $r(\cdot)$ has a positive intercept, a positive leading coefficient, and a positive root. Further, the positive leading coefficient and intercept implies there exists a negative root. Finally, any cubic with two distinct roots has a nonnegative discriminant, implying the discriminant must be nonnegative for large t_i. Hence, the limit of the discriminant is weakly positive. This derives a contradiction, as the limit of the discriminant, given that the polynomial converges to $1 + \kappa\alpha^2 - \alpha$ is $\kappa^2(1 - 4\kappa)$ using the cubic discriminant formula, which is strictly negative as we assumed $\kappa > \kappa^{c.k.} = 1/4$. □

If by contradiction $\theta(h^t) \to_p \theta$, then there exists an infinite subsequence $\{t_i\}_{i=1}^{\infty}$ of periods with revelation such that: $c_{3,t_i} + \kappa\alpha_{t_i}c_{4,t_i} \neq 0$, where such terms are defined in Equation (24). However, as $\theta(h^t) \to_p \theta$, $c_{5,t_i} \to_p 0$ (as Remark 3 implies $(K_{t_i})_{2,1} \to 0$), implying

$1 + c_{5,t_i} + \kappa\alpha_{t_i} c_{6,t_i} > 1$ (as $\kappa\alpha_{t_i} > 0$ and $c_{6,t_i} > 0$ by Remark 2) in the limit. Then, Equation (24) implies there are infinitely many noisy signals about v_t, implying $\mu(h^t) \to \mu$. As learning about θ, μ occurs, $|c_{3,t_i}|, |c_{4,t_i}|, |c_{5,t_i}|, |c_{6,t_i}| \to 0$. Therefore, Claim 1 holds deriving a contradiction with the assumption that there are infinitely many revealing periods.

Therefore, suppose $\mu(h^t) \to_p \mu$ but $\theta(h^t) \not\to_p \theta$. Therefore, there exists an infinite subsequence $\{t_i\}_{i=1}^{\infty}$ of periods with revelation such that: $1 + c_{5,t_i} + \kappa\alpha_{t_i} c_{6,t_i} \neq 0$, where such terms are defined in Equation (24). As $\mu(h^t) \to_p \mu$ but $\theta(h^t) \not\to_p \theta$, $\sigma^2_{\mu,t} \to 0$ and $\sigma^2_{\theta,t} \to \sigma^*_\theta$, because the variance is decreasing along the equilibrium path but bounded above zero as $\theta(h^t) \not\to_p \theta$. Further, $\theta(h^t) \not\to_p \theta$ implies that $\lim |c_{3,t_i} + \kappa\alpha_{t_i} c_{4,t_i}| = 0$ (this is the coefficient on s_t in the decision rule, else there would be infinitely many noisy signals about θ produced). By Remarks 1,2,3 we know $\lim |c_{4,t_i}| = \lim |c_{5,t_i}| = \lim |c_{6,t_i}| = 0$. Finally, α_{t_i} must be bounded by Lemma A3. Therefore, $\lim |c_{3,t_i} + \kappa\alpha_{t_i} c_{4,t_i}| = 0 \implies \lim |c_{3,t_i}| = 0$, and thus Claim 1 holds, similarly deriving a contradiction.

Proof of Proposition 2. What remains to show is that if $\kappa \leq \kappa^{c.k.}$, social learning about fundamentals and preferences occurs. The proof proceeds in three steps: (1) providing a sufficient condition for the signaling equilibrium to be revealing, (2) using this condition to show that the signaling equilibrium is revealing for infinitely many periods, and (3) that infinitely many periods of revelation implies social learning about fundamentals and preferences occurs.

Step 1: Fix beliefs, $\theta(h^t), \mu(h^t)$. By Lemma A2, it is sufficient to provide a condition whereby $\tilde{s}(\theta(h^t), \mu(h^t), \alpha, \kappa) \leq 1$ when $\alpha \cdot \kappa = 1/2$. This is sufficient because then when $\alpha \cdot \kappa = 1/2$, $(1 + \kappa\alpha^2) - \alpha \leq 0$, as $\kappa \leq \kappa^{\text{c.k.}} = 1/4$. Therefore, when $\alpha \cdot \kappa = 1/2$, $(1 + \kappa\alpha^2)\tilde{s}(\cdot) - \alpha \leq 0$, which is sufficient for the signaling equilibrium to be revealing by Lemma A2. Let

$$x_t = \gamma_t \mathbb{E}(\theta \mid s_t, v_t) + \frac{1}{2}\mathbb{E}(\mu \mid s_t, v_t),$$

so that y_t, as defined in Equation (24), equals $v_t + x_t$. Therefore,

$$\text{Var}(y_t) = \text{Var}(v_t) + 2\,\text{Cov}(v_t, x_t) + \text{Var}(x_t), \qquad \text{Cov}(v_t, y_t) = \text{Var}(v_t) + \text{Cov}(v_t, x_t),$$

and therefore given the definition of the sensitivity in Equation (25),

$$\tilde{s}(\cdot) = \frac{\text{Cov}(y_t, v_t)}{\text{Var}(y_t)} \leq 1 \iff \text{Cov}(v_t, y_t) \leq \text{Var}(y_t) \iff \text{Var}(x_t) + \text{Cov}(v_t, x_t) \geq 0. \tag{30}$$

Since $\text{Var}(x_t) \geq 0$, it suffices to provide a condition where $\text{Cov}(v_t, x_t) \geq 0$.

To compute $\text{Cov}(v_t, x_t)$, note that v_t is measurable with respect to (s_t, v_t), hence for any

integrable random variable W, $W = \mathbb{E}(W|s_t, v_t) + \zeta_t$ where $\text{Cov}(\zeta_t, v_t) = 0$. Applying this with $W = \theta$ and $W = \mu$ gives

$$0 \leq \text{Cov}(v_t, x_t) = \gamma_t \,\text{Cov}(v_t, \theta) + \frac{1}{2}\,\text{Cov}(v_t, \mu).$$

In words, this inequality states that v_t must be more predictive of $\frac{1}{2}\mu$ than it is of $\gamma_t\theta$. Using $v_t = \mu + \nu_t$ and independence of ν_t from (θ, μ),

$$\text{Cov}(v_t, \theta) = \text{Cov}(\mu, \theta) = \rho_t \sigma_{\theta,t} \sigma_{\mu,t}, \qquad \text{Cov}(v_t, \mu) = \text{Var}(\mu) = \sigma^2_{\mu,t}.$$

Therefore, a sufficient condition for the equilibrium to be revealing is

$$\gamma_t \rho_t \sigma_{\theta,t} \sigma_{\mu,t} + \frac{1}{2}\,\sigma^2_{\mu,t} \geq 0.$$

If $\rho_t \geq 0$ this expression is nonnegative for all $\gamma_t \geq 0$. Similarly if either $\sigma^2_{\mu,t}$ or $\sigma^2_{\theta,t}$ is zero then this condition always holds. Finally, if $\sigma^2_{\mu,t}, \sigma^2_{\theta,t}$ are non-zero and $\rho_t < 0$, then there is a cutoff on γ_t below which the equilibrium is revealing.

Step 2: If, by contrast, there was only finitely many periods of revelation, then there must exist a (possibly random) time T after which almost surely there exists no more periods of revelation. However, Step 1 showed that for any given beliefs, there is a positive probability that the equilibrium is revealing. Further, if period T is pooling, then the probability of revelation in period $T + 1$ is equal to that in period T as the beliefs do not change and γ_t is i.i.d.. Therefore the probability that the equilibrium is pooling in all subsequent periods is zero, proving claim 2.

Step 3: We now show that given infinitely many periods of revelation, social learning about preferences and fundamentals occurs. Denote by t_i the periods with revelation. Then the players observe $\beta_{s,t_i} s_{t_i} + \beta_{v,t_i} v_{t_i} = \beta_{s,t_i}\theta + \beta_{v,t_i}\mu + \mathcal{N}(0, \beta^2_{s,t_i} + \beta^2_{v,t_i})$. Suppose by contradiction social learning about θ and μ does not occur. As shown in Proposition A4 in the Online Appendix, there are two learning outcomes that can occur with infinitely many periods of revelation if social learning does not occur on both dimensions.

Case 1: Social learning about μ occurs but not θ. In this case, $\text{Cov}(v_t, x_t)$, as defined in Equation (30), is equal to $\gamma_t \rho_t \sigma_{\theta,t} \sigma_{\mu,t} + \frac{1}{2}\,\sigma^2_{\mu,t}$, converges to 0 as $\sigma^2_{\mu,t} \to 0$ and all other terms are bounded. Further, as social learning about μ occurs but not θ, then $x_t = \gamma_t \mathbb{E}(\theta \mid s_t, v_t) + \frac{1}{2}\mathbb{E}(\mu \mid s_t, v_t) = \gamma_t (K_t)_{1,1} s_t$, asymptotically. Finally, as learning about θ does not occur, $(K_t)_{1,1}$ is bounded above zero, implying the equilibrium, asymptotically, is revealing for any γ_t using the condition derived in Step 1. If this is the case, then in the limit the players observe $\gamma_t (K_t)_{1,1} s_t + v_t = \theta + \mu + \mathcal{N}(0, \gamma^2_t (K_t)^2_{1,1} + 1)$. However, as these errors are

independent, with finite variance, and the players learn μ, then the law of large numbers implies the players also learn θ, deriving a contradiction.

Case 2: Social learning about neither θ nor μ occurs. Then Proposition A4 implies that both (i) $\frac{\beta_{s,t_i}}{\beta_{v,t_i}}$ converges and (ii) $\rho_{t_i} \to -1$. Note $\beta_{s,t_i} = \gamma_{t_i}(K_{t_i})_{1,1} + \kappa\alpha_{t_i}(K_{t_i})_{1,2}$ is increasing in γ_{t_i} as Remark 1 implies $(K_{t_i})_{1,1} > 0$, as social learning about θ does not occur. Further, $\beta_{v,t_i} = 1 + \gamma_{t_i}(K_{t_i})_{2,1} + \kappa\alpha_{t_i}(K_{t_i})_{2,2}$, which is decreasing in γ_{t_i} as $(K_{t_i})_{2,1} < 0$ by Remark 3 as $\rho_{t_i} = -1$ and social learning about neither variable occurs. Therefore $\frac{\beta_{s,t_i}}{\beta_{v,t_i}}$ is non-constant in γ_{t_i}, which allows the players to separately identify θ, μ, contradicting the assumption that social learning fails. □

Proof of Proposition 3. The proof of the first statement follows from the second statement of Proposition 2. For the second statement, recall that Lemma 3 and Lemma A1 imply that an equilibrium is revealing if and only if $\inf_{\alpha\geq 0}(1+\kappa\alpha^2)\tilde{s}(\cdot) - \alpha \leq 0$. Therefore, for an equilibrium to be fragile, it must decrease the sensitivity. Lemma 3 implies that providing information about θ increases the sensitivity and thus precludes fragility of individual-oriented interventions. Further, Lemma 3 provides a condition where the derivative of the sensitivity is increasing in $\sigma^2_{\mu,t}$. That this condition is monotone in $\sigma^2_{\mu,t}$, implies that a discrete decrease in $\sigma^2_{\mu,t}$ (through a peer-oriented intervention), strictly decreases the sensitivity, which implies there exist histories that are fragile to peer-oriented interventions. □

Proof of Lemma 3. When the beliefs are uncorrelated, y_t, as defined in Equation (24) is

$$y_t = \frac{\gamma_t\sigma^2_{\theta,t}}{1+\sigma^2_{\theta,t}}s_t + (1+\frac{\kappa\alpha\sigma^2_{\mu,t}}{1+\sigma^2_{\mu,t}})v_t \implies$$

$$\tilde{s}(\cdot) = \frac{(1+\frac{\kappa\alpha\sigma^2_{\mu,t}}{1+\sigma^2_{\mu,t}})\text{Var}(v_t)}{(1+\frac{\kappa\alpha\sigma^2_{\mu,t}}{1+\sigma^2_{\mu,t}})^2\text{Var}(v_t) + (\frac{\gamma_t\sigma^2_{\theta,t}}{1+\sigma^2_{\theta,t}})^2\text{Var}(s_t)} = \frac{(1+\frac{\kappa\alpha\sigma^2_{\mu,t}}{1+\sigma^2_{\mu,t}})(1+\sigma^2_{\mu,t})}{(1+\frac{\kappa\alpha\sigma^2_{\mu,t}}{1+\sigma^2_{\mu,t}})^2(1+\sigma^2_{\mu,t}) + (\frac{\gamma_t\sigma^2_{\theta,t}}{1+\sigma^2_{\theta,t}})^2(1+\sigma^2_{\theta,t})}$$

Finally, one can take derivatives of this expression to verify the statement in the lemma. □

Proof of Proposition 4. Substituting $s_1 = \theta + \epsilon_1$ and $v_1 = \mu + \nu_1$ into $a_1 = \lambda_s s_1 + \lambda_v v_1$ gives

$$a_1 = \lambda_s\theta + \lambda_v\mu + u, \qquad u := \lambda_s\epsilon_1 + \lambda_v\nu_1, \tag{31}$$

where u is independent of (θ, μ) and satisfies $\text{Var}(u) = \lambda_s^2 + \lambda_v^2$. Conditioning on θ, we can rewrite this as

$$\frac{a_1 - \lambda_s\theta}{\lambda_v} = \mu + \frac{u}{\lambda_v}. \tag{32}$$

Since we assumed the initial beliefs are uncorrelated, the prior for μ conditional on θ remains $N(0, \sigma_\mu^2)$. Moreover, u/λ_v is independent Gaussian noise with variance

$$\operatorname{Var}\left(\frac{u}{\lambda_v}\right) = \frac{\lambda_s^2 + \lambda_v^2}{\lambda_v^2}. \tag{33}$$

Hence $(a_1 - \lambda_s\theta)/\lambda_v$ is a noisy signal of μ with independent noise, and standard Gaussian updating yields

$$\mathbb{E}(\mu \mid a_1, \theta) = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{\lambda_s^2+\lambda_v^2}{\lambda_v^2}} \cdot \frac{a_1 - \lambda_s\theta}{\lambda_v} \tag{34}$$

Finally, a^* is linear in a_1, by properties of the Gaussian distribution. Therefore,

$$\Delta(\theta) = \theta\left(1 - \frac{\sigma_\mu^2\,\lambda_s\lambda_v}{\lambda_s^2 + \lambda_v^2(\sigma_\mu^2 + 1)}\right) + c_\theta a_1, \tag{35}$$

for some constant c_θ. A symmetric calculation occurs for a peer-oriented intervention. □

Proof of Proposition 5. The proof directly follows from Proposition 6 when $\bar{g}(h^t) = 0$. □

The following lemma was not in the text, but is used to prove the subsequent propositions.

Lemma A4 *Let $f(\cdot)$ be a strictly increasing function and suppose $a(\cdot)$ satisfies*

$$a'(v) = \frac{\kappa f(v)}{v - a(v)}. \tag{36}$$

Then, there exists a solution to the above differential equation on $[v', +\infty)$ with $a(v') = v'$ if $\kappa \cdot \sup_{v \geq v'} f'(v) \leq \kappa^{c.k.}$, where v' is defined so that $f(v') = 0$. Further, no solution exists on $[v^*, +\infty)$ with initial condition $a(v^*) \leq v^*$ if $\kappa \cdot \inf_{v \geq v^*} f'(v) \geq \kappa^{c.k.}$.

Proof of Lemma A4. **Existence:** Since $f(v') = 0$ and f is C^1, for all $v \geq v'$ we have

$$f(v) = \int_{v'}^{v} f'(t)\,dt \leq (v - v') \sup_{v \geq v'} f'(v) \tag{37}$$

Define $\underline{d}(v) := (v - v')/2$. At any $v > v'$ with $v - a(v) = \underline{d}(v)$, combining (36) and (37) gives

$$\frac{d}{dv}\big(v - a(v)\big) = 1 - \frac{\kappa f(v)}{v - a(v)} \geq 1 - \frac{\kappa\,(v - v') \sup_{v \geq v'} f'(v)}{(v - v')/2} = 1 - 2\kappa \sup_{v \geq v'} f'(v) \geq \frac{1}{2} = \underline{d}'(v).$$

Thus $v - a(v)$ cannot cross below $\underline{d}(v)$ from above, so the solution starting at $v' + \epsilon$ satisfies

$$v - a(v) \;\geq\; \frac{v - v'}{2} \qquad \text{for all } v \geq v' + \epsilon. \tag{38}$$

In particular, the right-hand side of (36) is continuous and locally Lipschitz in a on the region $\{(v, a) : v > v',\, v - a \geq (v - v')/2\}$, so a unique C^1 solution exists on $[v' + \epsilon, \infty)$. Taking the limit as $\epsilon \to 0$ therefore proves existence.

Non-existence: Let $m := \inf_{v \geq v^*} f'(v) > 0$, then we have

$$f(v) \geq m(v - v^*) \implies a'(v) \geq \frac{\kappa m (v - v^*)}{v - a(v^*)}. \tag{39}$$

Therefore, there exists a v'' such that $a(v) > v^*$ for $v > v''$ as $\lim_{v \to \infty} a(\cdot) = \infty$. To see why, if $a(v)$ were bounded above, then $a'(\cdot)$ would be uniformly bounded above zero, contradicting $a(v)$ being bounded. Next note,

$$a'(v) = \frac{\kappa \frac{f(v)}{v - v^*}}{\frac{v - a(v)}{v - v^*}} = \frac{\kappa \frac{f(v)}{v - v^*} \frac{a(v) - v^*}{v - v^*}}{\frac{v - a(v)}{v - v^*}(1 - \frac{a(v) - v^*}{v - v^*})} \geq \frac{\kappa \frac{f(v)}{v - v^*} \frac{a(v) - v^*}{v - v^*}}{\kappa^{\text{c.k.}}} \geq \frac{m\kappa}{\kappa^{\text{c.k.}}} \frac{a(v) - v^*}{v - v^*}, \tag{40}$$

for $v > v''$ which follows as $x(1 - x) \leq 1/4$ for any $x \in [0, 1]$. In the above inequality, $x = (a(v) - v^*)/(v - v^*) \in (0, 1)$. Let $\bar{a}(\cdot)$ solve the differential equality with $\bar{a}(v'') = a(v'')$ obtained by making the above differential inequality tight. By the comparison theorem, $a(v) \geq \bar{a}(v)$ for all $v \geq v''$. Solving for $\bar{a}(\cdot)$ explicitly yields

$$\bar{a}(v) - v^* = \big(a(v'') - v^*\big) \left(\frac{v - v^*}{v'' - v^*} \right)^{\frac{m\kappa}{\kappa^{c.k.}}}. \tag{41}$$

Since v'' was chosen so that $a(v'') > v^*$, the prefactor in (41) is strictly positive. If $\kappa > \kappa^{c.k.}/m$, then $\frac{m\kappa}{\kappa^{c.k.}} > 1$, and hence there exists $\bar{v} > v''$ such that $\bar{a}(\bar{v}) > \bar{v}$. Thus $a(\bar{v}) \geq \bar{a}(\bar{v}) > \bar{v}$. However, this derives a contradiction as $a(v^*) \leq v^*$ and any solution to $a(\cdot)$ in Equation (36) cannot cross $a(v) = v$. □

Corollary A1 *Let $f(\cdot)$ be a strictly increasing function and suppose $a(\cdot)$ satisfies Equation (36). Then, there exists a solution to the above differential equation on $(-\infty, v']$ with $a(v') = v'$ if $\kappa \cdot \sup_{v \leq v'} f'(v) \leq \kappa^{c.k.}$, where v' is defined so that $f(v') = 0$. Further, no solution exists on $(-\infty, v^*]$ with initial condition $a(v^*) \leq v^*$ if $\kappa \cdot \inf_{v \leq v^*} f'(v) \geq \kappa^{c.k.}$ for any v^*.*

Proof of Corollary A1. Let $x := -v$ and define $\tilde{a}(x) := -a(-x)$ and $\tilde{f}(x) := -f(-x)$. For $x \geq -v'$ (equivalently $v \leq v'$), we have $\tilde{a}'(x) = a'(-x)$. Using the differential equation at

$v = -x$ gives,

$$\tilde{a}'(x) = \frac{\kappa \tilde{f}(x)}{x - \tilde{a}(x)}.$$

Applying Lemma A4 yields the stated conditions. □

Proof of Proposition 6. An immediate consequence of Bernheim (1994) is that the only equilibria satisfying D1 are central pooling equilibria with a derivative outside $[\underline{v}, \bar{v}]$ satisfying:

$$a_t'(v_t) = \frac{\kappa\big(v_t - g(v_t|h^t)\big)}{v_t - a_t(v_t)}\Bigg(\tag{42}$$

If κ is less than the threshold in the proposition, then denote by $v^*(h^t)$ the unique value where $v^*(h^t) = \mathbb{E}(\mu|h^t)$. Such a value exists by the Law of Iterated Expectation and it is unique as $g'(\cdot|h^t) < 1$. Lemma A4 and Corollary A1 imply the existence of a fully revealing equilibrium with $a_t(v^*(h^t)) = v^*(h^t)$. If κ exceeds the threshold in the proposition and if by contradiction either $\bar{v}$ (respectively, $\underline{v}$) is finite, then Lemma A4 (respectively, Corollary A1) implies that the differential equation associated with that equilibrium has no solution, deriving a contradiction. □

Proof of Corollary 1. In the Gaussian linear case, $g(\cdot) = \bar{\mu}(t) + \frac{\sigma^2_{\mu,t}}{\sigma^2_{\mu,t}+1}(v_t - \bar{\mu}(t))$. This implies $\underline{g} = \bar{g} = \frac{\sigma^2_{\mu,t}}{\sigma^2_{\mu,t}+1}$, which is increasing in $\sigma^2_{\mu,t}$. Further, the signaling equilibrium reveals v_t and every period of revelation strictly decreases the variance in the beliefs about μ. Therefore, a higher κ implies fewer periods of revelation, which proves the comparative statics on the beliefs. Finally, as noted in the text, the asymptotic adaptation loss and asymptotic utility are monotone with respect to the asymptotic variance of the beliefs. □

Proof of Proposition 7. An immediate consequence of Bernheim (1994) is that the only equilibria satisfying D1 are

$$a_t'(\tilde{v}_t) = \frac{\kappa m'(\tilde{v}|h^t)}{2(\tilde{v}_t - a_t(\tilde{v}))}. \tag{43}$$

Further, since $\lim_{x\to\infty} m(\cdot|h^t) = \lim_{x\to-\infty} m(\cdot|h^t) = \infty$, there must exist a point where $m'(\cdot|h^t) = 0$. Now, using an identical argument to that in Proposition 6, the result in this proposition directly follows from Lemma A4 and Corollary A1. □

Proof of Corollary 2. In the Gaussian analysis, $m(\cdot)$ takes the following closed form:

$$m(\tilde{v}_t \mid h^t) = \frac{\sigma_v^2 \cdot \frac{\sigma_{\theta,t}^4}{1+\sigma_{\theta,t}^2}}{\sigma_v^2 + \frac{\sigma_{\theta,t}^4}{1+\sigma_{\theta,t}^2}} + \left(\frac{\sigma_v^2}{\sigma_v^2 + \frac{\sigma_{\theta,t}^4}{1+\sigma_{\theta,t}^2}}\right)^2 (\tilde{v}_t - \bar{\theta}_t)^2. \tag{44}$$

The results then follow from noting that the second derivative is constant with respect to $\tilde{v}_t$, independent of $\bar{\theta}(t)$, and that this second derivative is monotone decreasing in $\sigma_{\theta,t}^2$. Further, $\sigma_{\theta,t}^2$ decreases following each period of revelation, so that a higher κ implies weakly fewer periods of revelation, proving the comparative statics in the corollary. □