Conformity Concerns: A Dynamic Perspective

Roi Orzach*

November 1, 2024

Please click here for the latest version.

Abstract

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

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

*Department of Economics, MIT, 50 Memorial Drive, Cambridge MA 02142 (e-mail: orzach@mit.edu).

I thank Alessandro Bonatti, Glenn Ellison, Robert Gibbons, and Stephen Morris for countless conversations and additionally thank Chiara Aina, Charles Angelucci, Ian Ball, Abhijit Banerjee, Inga Deimen, Stefano DellaVigna, Tore Ellingsen, Drew Fudenberg, Ying Gao, Saumitra Jha, Jackson Meija, Navin Kartik, Simone Meraglia, Paul-Henri Moisson, Ellen Muir, Vishan Nigam, Shelli Orzach, Jacopo Perego, Nicola Persico, Kramer Quist, Daniel Rappoport, Rafaella Sadun, Tomasz Sadzik, Lones Smith, Cass Sunstein, Jean Tirole, Alexander Wolitzky, as well as participants at the MIT Organizational Economics and Theory groups, 2024 North American Summer Meeting of the Econometric Society, and 2024 Stony Brook Game Theory Conference for helpful comments.

1 Introduction

Substance abuse is an important public health concern and greatly contributes to mortality rates across the world (cf. Sheikh et al., 2018). For decades public health officials educated adolescents on the cost of substance abuse, but these "initial attempts at prevention were ineffective because they focused primarily on lecturing students about the dangers and long-term health consequences of substance use" (Griffin and Botvin, 2010). These attempts failed because they assumed an "individual-oriented" (Prentice and Miller, 1993) cognitive model about their participants: individuals' decisions about substance use were imagined to be based purely on anticipated health outcomes for the individual. In contrast, informing students about their peers' true preferences towards substance use was much more effective: Schroeder and Prentice (1998) found that undergraduates who learned their peers' preferences towards alcohol reduced their consumption by 40 percent compared to those who were informed about just the health costs. Schroeder and Prentice (1998) call this a "peer-oriented" cognitive model.

Public health research rationalized this finding by noting that many individuals overestimate how much their peers enjoy alcohol (cf. Prentice and Miller, 1993). This misperception, combined with a desire to conform, motivated people to partake in substance abuse. Such phenomena are pervasive. Political endorsements (Loury, 1994; Geiger and Swim, 2016), female labor force decisions (Bursztyn et al., 2020a), and corporate board decisions (Westphal and Bednar, 2005; Chang et al., 2019) are interpreted as influenced in similar ways. Across these examples, inefficiencies may arise from two sources: misperceptions of the private benefits of such decisions and misperceptions of peers' attitudes towards these decisions.

In this paper, I provide a general framework that explains how these two misperceptions interact, why they hence may both persist, and why some interventions may be more successful than others. In the model, a community of agents attempts to learn two initially unknown variables: a fundamental state (which could be the health costs of substance abuse) and their peers' average preferences. Learning occurs both privately and socially. Privately, each individual receives a signal about the fundamental state and his own preferences, which are informative about his peers' preferences because they are drawn from the same population distribution. Further, the players may be able to learn the unknown variables by observing the decisions of their peers. 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? In answering this question, I depart from the "individual-oriented" model and explicitly model the desire to conform. This approach generates new predictions regarding the success or failure of social learning. Additionally, in cases where social learning fails, it provides novel insights into optimal interventions.

Formally, each individual seeks to maximize a utility function that combines their private utility with a conformist utility component. Their private utility is determined by how well their decision adapts to their private preferences and the fundamental state. Their conformist utility, on the other hand, is determined by the exogenous conformity concerns (a parameter common to all individuals) multiplied by the gap between how the community perceives their type and the true average preferences of the public. While the exogenous conformity concerns is constant over time, the endogenous reputational penalty from choosing decisions different from one's peers' decisions changes as the public beliefs about the fundamental state and average preferences become more precise. I refer to this endogenous reputational penalty as the "effective conformity concerns."

The key mechanism potentially preventing social learning is that once the effective conformity concerns are sufficiently large, an individual is unable to adapt to his private preferences without incurring a large reputational penalty. Further, because adaptation to a player's private signal about the fundamental state might be attributed to a player's preferences, such adaptations are similarly discouraged. Therefore, when the effective conformity concerns are large enough each decision is independent of a player's private information. This independence implies that no new information is publicly learned in the current period, resulting in an identical situation for the next period, and, by induction, for all subsequent periods. As a result, if the effective conformity concerns ever become sufficiently large, all subsequent players will pool on inaccurate perceptions of the fundamental state and of the average preferences.

Finally, interventions to improve efficiency are needed primarily when social learning fails. Because social learning fails when the effective conformity concerns are large, then when interventions are needed, they should target conformity concerns (via information about average preferences) as opposed to adaptation loss (via information about the fundamental state).

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 private preferences, as described above. In this benchmark, exogenous conformity concerns are still a parameter shared by all individuals, however in this static benchmark there is no learning, implying that the effective conformity concerns are equal to the exogenous conformity concerns. In this scenario, conformity concerns place a penalty on the degree to which players adapt to their private preferences (cf. Bernheim, 1994). I show that all players choose the same decision independent of their private preferences if and only if the conformity concerns exceed a given threshold (Proposition 1).[1] This benchmark shows how, when conformity concerns are significant, an individual's decisions may cease to reflect their true preferences.

In Section 4, I analyze the complete model where conformity concerns interact with uncertainty about the fundamental state and average preferences. The necessary and sufficient condition for "social learning to occur" (i.e., for the players to learn asymptotically the fundamental state and the preferences of their peers) is that there exist infinitely many periods of a decision rule with revelation. This condition requires that the decision rule involves revelation when the beliefs about the fundamental state and average preferences are arbitrarily precise. As a result, a necessary condition for social learning to occur is that the conformity concerns must be less than the threshold for revelation in the benchmark (Lemma 2).

Unsurprisingly, when exogenous conformity concerns are sufficiently small, this places an upper bound on the effective conformity concerns. Consequently, if exogenous conformity concerns are sufficiently small, learning occurs regardless of prior beliefs. Finally, I show that when the exogenous conformity concerns are intermediate, the learning outcome depends on the initial beliefs (Proposition 2).

When exogenous conformity concerns are such that learning fails in the limit, the players face three inefficiencies. First, the players are unable to adapt to their private preferences. Further, the players' decisions are based on imprecise beliefs about *both* the fundamental state and the average preferences of their peers.

Given these two imprecise beliefs, Section 5 analyzes the different effects of *peer-oriented* interventions, which inform players about their peers' preferences (e.g., how "cool" substance abuse is thought to be by others), versus *individual-oriented* interventions, which inform players about the fundamental state (e.g., the health costs of substance abuse).[2] If individuals had no desire for conformity, they would disregard their peers' preferences entirely. In such a scenario, an individual-oriented intervention would always be the preferred approach. However, in such situations interventions are, arguably, not needed; if there were no desire for conformity-then, asymptotically, the players would eventually learn the fundamental state. For this reason, I compare the effect of peer-oriented and individual-oriented interventions based on their ability to break a pooling equilibrium when the exogenous

[1] In Bernheim (1994), regardless of the strength of conformity concerns, there always exists an equilibrium with partial revelation. In my model, when conformity concerns are sufficiently high, the only equilibrium will be fully pooling. I discuss this difference more in Section 6.2.

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

conformity concerns are high.

When exogenous conformity concerns are high, I show that if the beliefs about the fundamental state and average preferences are uncorrelated, then individual-oriented interventions can never break a pooling equilibrium. Intuitively, players may deviate from past decisions for one of two reasons: first, if a player's signal implies the fundamental state differs from the public beliefs; second, if a player's private preferences motivate a different decision from his peers' decisions. Importantly, because deviations are only partially ascribed to a player's preferences, imprecise beliefs about the fundamental state encourage adaptation, implying that individual-oriented interventions increase the effective conformity concerns and will not break a pooling decision rule.

In contrast, peer-oriented interventions may break a pooling outcome. There are two competing effects of a peer-oriented intervention. First, when players are unsure what the average preferences of the population are, their own preferences serve as an informative signal. Given that a player wants the public perception of his preferences to be close to the average preferences, choosing a decision that is responsive to his own private preferences is optimal; hence, a non-pooling equilibrium will exist. The competing force is that if a player is unsure what the average preferences of the population are, then the population is also uncertain of that player's preferences. The latter uncertainty implies that the population's inference about that player's preferences places a greater weight on his decision rather than the prior. I show that if the uncertainty about the fundamental state is large, the competing effect is stronger, and thus peer-oriented interventions may break a herd. Generalizing the definition in Bikhchandani et al. (2021), the pooling equilibrium is "fragile" to peer-oriented interventions but not to individual-oriented interventions (Proposition 4). Moreover, even when neither intervention can break a pooling equilibrium, the peer-oriented intervention may have a more substantial impact on the decision players pool on (Proposition 5). This is because a peer-oriented intervention enables players to pool on a more efficient decision.

My findings are consistent with several key studies (e.g., Schroeder and Prentice, 1998; Bursztyn et al., 2020a; Gulesci et al., 2023) advocating for peer-oriented interventions in situations with significant conformity concerns. Furthermore, my research provides a mechanism supporting Sunstein's (2019) argument that sharing information about others' preferences can lead to the revelation of true beliefs and preferences. As Sunstein notes, this process "unleash[es]" people, allowing them to "reveal what they believe and prefer," ultimately facilitating the discovery of "preexisting beliefs, preferences, and values."

Further, in my model, before a peer-oriented intervention, the players will have incorrect beliefs about the true average preferences of the group. This misperception is referred to as "pluralistic ignorance" in the social psychology literature, which is summarized below in

Section 1.1. This literature notes that a primary source of pluralistic ignorance is a failure to recognize and adapt to changes to the groups' preferences. Motivated by this finding, I consider an extension where the underlying preferences of the population change over time and show that conformity concerns exacerbate the lag between the public beliefs and true preferences (Proposition 6).

Finally, the paper explores three further extensions that provide robustness checks for the main analysis. These extensions show that the qualitative features of the analysis do not depend on the following assumptions: short-run players (Proposition 7), linear decision rules (Proposition 8), or Gaussian random variables (Proposition 9).

As I elaborate below, this paper provides a new rationale for why social learning fails: conformity concerns. When conformity concerns are sufficiently small, the results from the earlier literature continue to hold and discrete actions or boundedly informative signals may prevent social learning. However, for higher (but, importantly, non-infinite) values of the conformity concerns social learning will fail due to the conformity concerns. In such settings, my model delivers predictions consistent with the empirical literature about effective interventions, whereas the social learning literature without conformity concerns would conjecture that individual-oriented interventions are always optimal.

1.1 Related Literature

This paper is related to three strands of literature: models of social learning, models of decision-making with reputational concerns, and the empirical literature about pluralistic ignorance and interventions. After presenting my results, I will connect my findings to the empirical literature in Section 4.3 and Section 5.

This paper is closely related to the literature on social learning. In this literature, social learning is the process in which players learn about their environment through their peers' decisions. Social learning is said to occur when asymptotically the players make optimal decisions. For instance, Banerjee (1992) and Bikhchandani et al. (1992) study models where players sequentially receive information and make a decision. If a sequence of past decisions is informative about the signals those players received, then players may rationally choose to stop utilizing their own signals, causing social learning to stop. In contrast, for the quadratic-loss environment I consider Lee (1993) shows that continuous decisions are a sufficient condition for social learning to occur.[3] Further, an implication of Kartik et al. (2024) is that with quadratic-loss payoffs, "directionally unbounded beliefs," which the

[3] More generally, Ali (2018) shows that if preferences are "responsive," defined as every distinct belief having a distinct optimal action, then social learning occurs.

Normal distribution satisfies, is sufficient for social learning.[4] I allow for both responsive decisions and directionally unbounded beliefs, and yet find that social learning can fail in the presence of conformity concerns. In contrast, the social learning literature has documented other obstacles to social learning such as costs of acquiring information (cf. Burguet and Vives, 2000; Chandrasekhar et al., 2018), misspecified priors (cf. Bohren, 2016; Frick et al., 2020), non-bayesian updating (cf. Golub and Jackson, 2010), changing fundamentals (cf. Dasaratha et al., 2023), or differential observability assumptions (cf. Banerjee and Fudenberg, 2004; Arieli and Mueller-Frank, 2019).

Further, this paper relates to the literature on reputational concerns. This literature typically analyzes settings where an agent takes an observable decision attempting to both (i) adapt the decision to a signal and (ii) make the observer think the agent is a "good type". In Scharfstein and Stein (1990), an agent wants to be perceived as competent, and in Morris (2001), an agent wants to be perceived as un-biased.[5] In both these papers, and most of the literature on reputational concerns, there is a single observer viewing the player's decision and the preferences of this observer are common knowledge.[6] In contrast, I focus on environments where the decision-maker has multiple observers and the ideal perceived type for the decision-maker is different for each potential observer, implying that the sender has single-peaked preferences over the receiver's beliefs. A subset of the reputation literature analyzes such preferences: 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 and modeled in Bernheim (1994). However, in all these papers, the interaction is static and the lone decision-maker maximizes over the distribution of observers. Instead, in my dynamic analysis there will exist aggregate uncertainty over the distribution of observers that will be partially resolved as the game unfolds. When this uncertainty is not fully resolved, the players conform to incorrect perceptions of their peers, a finding not present in the literature on reputational concerns.

There is a literature that considers the effect of reputation and coordination on social

[4]Relatedly, Smith and Sørensen (2000) show that if preferences are monotone and the state of the world is binary a non-zero probability of an arbitrarily precise signal is sufficient for efficient social learning to occur.

[5]Braghieri (2021) provides empirical support for the prediction in Morris (2001) that political correctness may render speech uninformative.

[6]There are many papers which analyze different settings, but have the same feature: there is a known type the agent wants to be perceived as. See Canes-Wrone et al. (2001); Ely and Välimäki (2003); Ottaviani and Sørensen (2006); Braghieri (2021); Rappoport (2022). Ely and Välimäki (2003) is especially relevant as Ely and Välimäki (2003) show that when players are sufficiently patient, the unique equilibrium is a pooling equilibrium. I show a similar result in Subsection 6.1, however the mechanisms are different.

learning. For instance, Smith et al. (2021) shows contrarianism behavior should be rewarded for efficient informational herding. With conformity concerns, not only is contrarianism not rewarded, it is actively punished, resulting in inefficient learning. Additionally, Angeletos et al. (2007) considers an environment where players may receive benefits from coordinating on similar decisions. These are distinct from conformity concerns: a player's preferences over his perceived preferences.[7] This distinction results in different predictions for asymptotic learning and allows for the possibility of peer-oriented interventions.

The impact of conformity concerns on social learning is considered in Li and Van den Steen (2021) and Fernández-Duque (2022). In both models, players can either publicly support or oppose an issue. An individual's support for a particular decision is influenced by two factors: their personal preferences and their perception of social approval. These models are different from my work because in those papers (i) there is uncertainty about only the players' preferences, not the fundamental state of the world, and (ii) decisions and preferences are binary. Distinction (i) allows my work to discuss when an individual-oriented or a peer-oriented intervention is preferred. Distinction (ii) allows my work to predict a failure in social learning where previous work would not because the social learning literature without conformity concerns already predicts failures in social learning with discrete decisions. However, with conformity concerns, herding occurs with continuous decision-making. To see why, note that even a mild adaptation to one's private information may be met with an extreme change in reputation, preventing players from utilizing the continuous decision set.

Finally, this paper also contributes to a century of literature on "pluralistic ignorance," which refers to the systematic misperception of peers' preferences. For a review within economics see Bursztyn and Yang (2022) and within social psychology see Miller (2023). This literature suggests that peer-oriented interventions (i.e., ones that correct pluralistic ignorance) are preferred in a variety of settings: substance abuse (cf. Schroeder and Prentice, 1998), female labor force participation (cf. Bursztyn et al., 2020a), and religious norms (cf. Gulesci et al., 2023). My paper provides a general framework that develops a prediction that is consistent with these applied literatures: pluralistic ignorance arises and peer-oriented interventions are needed when conformity concerns are high.

The rest of the paper is organized as follows. Section 2 describes the model, equilibrium assumptions, and defines relevant terms. Section 3 analyzes a benchmark of the model with

[7] Bernheim (1994) provides a discussion of the difference between "popularity" concerns where each player's preferred decision is similar to his peer's decision and "conformity" concerns where each player's preferred decision is shaped by his perceived preferences. Bernheim (1994) shows popularity concerns yield vastly different equilibria than conformity concerns: for instance, popularity concerns will result in strictly monotone decision rules whereas conformity concerns result in pooling. Importantly, for popularity concerns there should be no difference between private and public decisions, whereas in Bursztyn et al. (2019, 2020b); Braghieri (2021), and many others we see differences.

common knowledge. Section 4 contains the main analysis. Section 5 extends the model to include interventions and additional determinants of pluralistic ignorance. Section 6 contains the three robustness extensions. Finally, Section 7 concludes and the Appendix contains proofs for all statements not shown in the text.

2 Model

Consider a community whereby, each period, a player makes a decision attempting to adapt to his private information and private preferences while possessing conformity concerns. The first subsection describes this set-up and the second subsection discusses the equilibrium refinements used throughout the analysis.

2.1 Model Description

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 both a common fundamental state, $\theta \sim N(0, \tau_\theta)$, and the average preferences of the group, $\mu \sim N(0, \tau_\mu)$.[8] Each player's private information includes a private signal $s_t = \theta + \epsilon_t$ and his private preferences $v_t = \mu + \nu_t$. I assume ϵ_t and ν_t are independent within and across periods, and that both are Gaussian random variables with a mean normalized to zero and a variance 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 private preferences. The weight of the fundamental state, $\gamma_t \geq 0$, is publicly observable and discussed further below. Second, while each player observes his own preferences, the player prefers the public's perception of his preferences to be close to the average within the community, which represents the conformity concerns.[9] Define by $\phi(b|h_t, a_t)$ the probability distribution over preferences b that take decision a_t given the history of previous decisions.[10] Further, for now, $\phi(\cdot)$ is

[8] The Gaussian assumption allows for clean comparative statics and to solve the model under uncertainty about both the preferences of others and the fundamental state. The benchmark in Section 3 where θ, v are common knowledge holds for any distribution of v_t, as discussed in Section 6.2. Section 6.3 discusses generalizations of the learning framework to other distributions.

[9] I show in Proposition 10 in Appendix B that one would get qualitatively similar results if, throughout the analysis, each player wanted his perceived preferences to be b units higher than μ.

[10] There are two distinct reasons why previous decisions impact the inference function. First, previous decisions impact the equilibrium perception about θ, and thus a player may choose a higher decision due to a higher perception of θ rather than higher preferences. Second, previous decisions impact the equilibrium perception of the average preferences of the population, which impact the

unconstrained off-path. The total utility for player t is thus,

$$u_t(a_t; v_t, s_t|h_t) := -\mathbf{E}_{\theta,\mu}\Big((a_t - \gamma_t\theta - v_t)^2 - \kappa \int (b-\mu)^2 \phi(b|h_t, a_t)db \Big| v_t, s_t, h_t\Big). \quad (1)$$

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 preferences. The second term is the conformity term, scaled by $\kappa \geq 0$. The term within the parenthesis is a reduced-form representation of conformity: player t wants the community's perception of his preferences, b, to be close to the true average preferences of the community, μ. Given that the loss function is quadratic, one can show that from the stand-point of which decisions are taken, this is equivalent to player t preferring that such an inference be close to that of a randomly drawn member of the community.[11]

Information: I assume that the public history takes the form $h_t = \{\gamma_1, a_1, \ldots, \gamma_{t-1}, a_{t-1}, \gamma_t\}$. This assumption states that there is full observability of the decisions and when they were made. Given this assumption, one can interpret the γ_t's as commonly observed time fixed-effects determining whether the fundamental state or one's preferences are comparatively more important. These time fixed effects are modeled as (i) $\gamma_t \leq 1 \; \forall \; t$, and (ii) there being an $m \in \mathbf{N}$ such that $\gamma_{t+m} = \gamma_t$ and $\gamma_{t+1} \neq \gamma_t$. These assumptions are only necessary in the analysis with uncertainty about θ and μ. Without these assumptions there is only one "moment condition" for the players to separately infer θ and μ, potentially resulting in incomplete learning. However, in Section 9.4.4, I show that such learning outcomes are not locally stable, and this assumption removes their existence.[12]

I analyze Perfect Bayesian Equilibria (cf. Fudenberg and Tirole, 1991) satisfying the following requirements each period, and I will discuss each requirement below.

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

2. *Social Optimality:* The players always play the linear equilibrium that maximizes total surplus.

equilibrium perception of a given player's preferences.

[11]Both interpretations correspond to Bernheim (1994) when the uncertainty over preferences is degenerate.

[12]These learning outcomes resemble the confounded learning outcomes in Smith and Sørensen (2000). Unlike Smith and Sørensen (2000), these outcomes are not locally stable. Intuitively, what this assumption states is that players may place different weights on the fundamental state, e.g., at the end of the semester with respect to drinking or during election cycles in political speech examples.

2.2 Equilibrium Selection:

Linearity: The restriction to linear equilibria is common when studying the normal learning model as it allows for greater tractability. Generally, this assumption is with loss of generality. Despite this, I show in Section 6.2 that when the distribution of preferences has full support and there is no uncertainty over θ (the fundamental state) or μ (the average preferences in the population), the only equilibria with revelation which satisfy D1 from Cho and Kreps (1987) are the linear equilibria. Upon adding uncertainty over θ, through signals s_t, one must add distributional assumptions to ensure linearity is without loss. To see why, note that a_t will be a function of $v_t + \gamma_t \mathbf{E}(\theta|s_t, v_t)$, and thus the public inference of v_t will be based on the sum $v_t + \gamma_t \mathbf{E}(\theta|s_t, v_t)$. If this inference is non-linear, then generally the decision rule will be non-linear. In the Gaussian case, this inference is linear, and in Appendix B, I prove that the only equilibria with revelation that satisfy D1 remain linear. Similarly, upon adding uncertainty over μ, if player t's inference about μ is non-linear in his private preferences, v_t, then generally the equilibrium will be non-linear. Again, the Gaussian assumption ensures that the inference is linear, and in Appendix B, I show that this assumption implies that the only equilibria satisfying D1 with revelation are linear.

Social Optimality: I assume that for each period, t, the equilibrium decision rule in period t maximizes the expected surplus out of all linear decision rules in period t.[13] This refinement is identical to one where the sequence of equilbria across periods maximizes the discounted expected surplus of the players from the class of linear equilibria in all periods. This refinement implies the players utilize a decision rule with revelation whenever one exists. Intuitively, players always prefer the linear equilibrium with revelation to pooling on any decision, a^*. In the revealing equilibrium, each type can always choose a^*, and that they do not implies they have a weak preference to not. Further, the conformity loss is independent of the equilibrium chosen by Bayes Plausibility.

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

3 Common Knowledge Benchmark

This section analyzes the impact of conformity on decision-making and mutes uncertainty about the fundamental state and the preferences of others. To do so, I assume θ and

[13] Without the social optimality refinement, then for any sequence of values $x_1, \ldots,$ there exists an equilibrium where all players pool on x_t in period t. Such equilibria are removed by this requirement.

μ are common knowledge, and, without loss of generality, are both equal to zero.[14] Further, without uncertainty, there is no time dependence, thus it is without loss of generality to consider the decision rule of a player with type v. A signaling equilibrium requires that this decision rule is linear in v. Given linearity, there are two cases: a constant and a strictly increasing decision rule. If the decision rule is constant, then this decision rule is defined as "fully-pooling." Further, for any fully-pooling decision rule, there exist off-path beliefs that deter any deviations from the pooling decision.

Any decision rule, $a(v)$ that is non-constant is defined as "revealing" and is determined by $\hat{v}(a) = \alpha a + \beta$, which was defined as the posterior expectation of a player's type given his decision. The necessary and sufficient condition for these beliefs to constitute an equilibrium is that given $\hat{v}(a)$, the decision rule that maximizes a players utility, $a(v)$, must result in a consistent conjecture of $\hat{v}(a)$. The first-order condition for the decision rule given a conjecture $\hat{v}(a) = \alpha a + \beta$ is:

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

Further, these beliefs constitute an equilibrium if and only if:

$$1 + \kappa\alpha^2 = \alpha \text{ and } \beta = \kappa\alpha\beta. \tag{3}$$

It is immediate that $\beta = 0$ is a solution to the latter equality, and, further, it is the unique solution for any α that solves the former. While one can solve the former equality, Figure 1 provides intuition why an equilibrium with revelation cannot exist for high values of conformity concerns. Figure 1 depicts as a function of the conjectured slope of the beliefs, α, the resulting beliefs from the best response, $1 + \kappa\alpha^2$, for two different values of κ. Further, note that Figure 1 also provides intuition for why the fully-pooling decision rule is an equilibrium. In the pooling equilibrium, the slope of the decision rule is zero, implying that the slope of the beliefs (i.e., the inverse of the decision rule) is infinity. Finally, the best response to such beliefs is to choose the pooling decision, which will generate such beliefs in equilibrium.

[14] As a result of θ being common knowledge, the players disregard their signals s_t, which removes any need for distributional assumptions about the signals. This simplified analysis is similar to that of Bernheim (1994). Section 6.2 examines this environment without assuming linearity and offers a comprehensive comparison with the findings of Bernheim (1994).

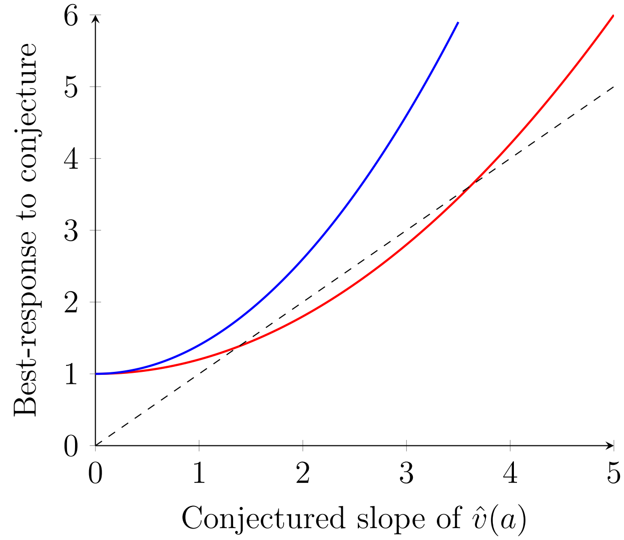

Figure 1: Existence of Linear Equilibria with Revelation

The x-axis represents a conjectured slope of the posterior beliefs of the community given a decision, $\hat{v}(a)$. The y-axis depicts the beliefs that result from the best response of the players to such a conjecture as stated in Equation (3). In Blue is the best response when the conformity concerns are high ($\kappa = .4$) and in Red the best response when the conformity concerns are low ($\kappa = .2$).

Figure 1 provides complementary reasons why no equilibrium with revelation exists for high values of κ. Intuitively, (i) as the conformity concerns increase, the players must conform more resulting in a lesser slope of $a(v)$. As $\hat{v}(a)$ is the inverse of $a(v)$, then the slope of $\hat{v}(a)$ is high. Further, (ii) the players cannot conform too much while still maintaining an equilibrium with revelation. If this occurred, then the slope of $a(v)$ must be small, implying a large slope of $\hat{v}(a)$. If the conjectured slope is too large, the players have a large incentive to conform as their decision greatly impacts their reputation. Combining these intuitions, if the conformity concerns are high then (i) dictates that the slope of $\hat{v}(a)$ must be large, but (ii) dictates that such slopes cannot constitute an equilibrium.

Finally, when the conformity concerns are low enough that an equilibrium with revelation exists, then three different linear Perfect Bayesian Equilibria exist. As the conformity loss is fixed across the equilibria and equal to κ multiplied by the variance of v, we can focus on the adaptation loss. Further, a lesser slope of $\hat{v}(a)$ corresponds to a greater slope of $a(v)$, implying that the equilibrium with the lowest slope of $\hat{v}(a)$ results in the best adaptation loss. Therefore, the social optimality refinement pins down a unique signalling equilibrium.

Proposition 1 (**Commonly Known Environment**)

There exists a threshold value, $\kappa^{c.k.}$, such that the unique signaling equilibrium as defined in Section 2.2 is characterized by the following decision rule:

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

where κ denotes the weight on conformity. Given such a decision rule, equilibrium utility is

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

This proposition summarizes the intuitions from the figure. First, the degree to which players adapt to their preferences is decreasing with respect to the conformity concerns, κ. To see this, note that when $\kappa = 0$, $a(v) = v$ and as κ increases up to the threshold $\kappa^{c.k.}$, $a(v) = v/2$. The reason for this decrease is that each player has an added incentive to conform when the conformity concerns increase. Finally, when the incentive to conform becomes sufficiently high, i.e., $\kappa > \kappa^{c.k.}$, there is no decision rule with revelation and the signaling equilibrium is fully pooling. Importantly, such fully-pooling decision rules provide no information about a player's type, and this observation will be key in the main analysis.

4 Analysis

This section begins with an analysis of the complete model where the players learn both the fundamental state and the preferences of their peers while possessing conformity concerns. I show that the players learn the fundamental state if and only if they learn the preferences of their peers. Further, such learning occurs whenever the conformity concerns are sufficiently high and fails whenever the conformity concerns are sufficiently low. The second subsection present additional comparative statics about the asymptotic utility and the asymptotic precision of the beliefs about the fundamental state and show that both of these terms are decreasing in κ. In doing so, I assume away uncertainty about the population's average preferences and analyzes how the players learn the fundamental state.[15] Incorporating both dimensions of uncertainty allows for predictions that neither uni-dimensional learning model will produce. However, doing so complicates the analysis by

[15] Further, in Subsection 5.5, I conduct an analysis where the fundamental state is common knowledge, but the average preferences of the community is uncertain. In that analysis, conformity concerns prevent the players from learning the preferences of their peers, and provides similar comparative statics.

requiring a joint update in the posterior beliefs each period. In the final subsection, I relate the findings of my baseline theoretical model to the empirical and qualitative literatures discussed in Section 1.1.

Before delving into the analysis, I define notation that appears throughout. Given the Gaussian set-up, the joint distribution of the public beliefs about θ and μ at any time t remain jointly Gaussian.[16] Define $\theta(t) = \mathbf{E}(\theta|h_t)$ and $\mu(t) = \mathbf{E}(\mu|h_t)$. These random variables have the following joint distribution.

$$\begin{pmatrix} \theta(t) \\ \mu(t) \end{pmatrix} \sim N\left(\begin{pmatrix} \bar{\theta}(t) \\ \bar{\mu}(t) \end{pmatrix}, \begin{pmatrix} \tau_{\theta,t} & \rho_t\sqrt{\tau_{\theta,t}\tau_{\mu,t}} \\ \rho_t\sqrt{\tau_{\theta,t}\tau_{\mu,t}} & \tau_{\mu,t} \end{pmatrix}^{-1} \right) \tag{6}$$

As is common, it will be convenient to work with the precision matrix, defined as the inverse of the variance matrix in the above equation. Below, I introduce definitions.

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

In the signaling equilibrium, the inference function, $\phi(b, h_t, a_t)$, will be Gaussian, the mean of which will be denoted as $\hat{v}(a_t) = \int b \cdot \phi(b, h_t, a_t)$. Within the quadratic set-up the variance of $\phi(b, h_t, a_t)$ and the variance of μ are independent of the decision taken. These observations allow for the following simplification of the utility function up to some constant c_t as stated below.

$$\begin{aligned} u_t(a_t; v_t, s_t|\theta(t), \mu(t), \gamma_t) :=& -\mathbf{E}_{\theta,\mu}\Big((a_t - \gamma_t\theta - v_t)^2\Big|\theta(t), \mu(t), v_t, s_t\Big) \\ &- \kappa\mathbf{E}_{\theta,\mu}\Big(\big(\hat{v}_t(a_t) - \mathbf{E}(\mu)\big)^2\Big|\theta(t), \mu(t), v_t, s_t\Big) - c_t. \end{aligned} \tag{7}$$

The first term is the adaptation loss. The second term is the squared difference from the mean perception of player t and the average preferences of the population, and the final term is the variance stemming from the uncertainty over μ and θ. Given this utility, I define the following.

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

[16] Notably, while the prior beliefs about θ and μ in period 1 are independent, in any subsequent period, the beliefs about θ and μ are dependent as both condition on the same sets of decisions.

Throughout I provide results discussing both notions of asymptotic efficiency. However, the comparative statics for one notion of asymptotic efficiency coincide with the comparative statics for the other because the difference in these terms is the expected conformity loss which is pinned down by Bayes plausibility.

4.1 Determinants of Social Learning

This subsection analyzes the complete model where the players attempt to socially learn θ and μ as described in Section 2.1. To analyze this environment, I consider an arbitrary period, t. Recall that the beliefs about θ and μ follow a bivariate normal distribution. These beliefs are sufficient statistics for the sequence of past decisions, $a_1, \ldots, a_{t-1}$, and the relative weights on the fundamental state in each period $\gamma_1, \ldots, \gamma_{t-1}$ in a linear equilibrium.

To analyze the signaling equilibrium, fix any equilibrium that is not fully pooling and a conjecture for $\hat{v}_t(a_t) = \alpha_t a_t + \beta_t$, where the consistent conjectures for α_t and β_t will, of course, depend on the beliefs at period t. In this equilibrium, player t's first-order condition for his decision rule is:

$$a_t(1+\kappa\alpha_t^2) = \gamma_t \mathbf{E}\big(\theta \mid \theta(t), \mu(t), s_t, v_t\big) + v_t + \kappa\alpha_t\beta_t + \kappa\alpha_t \mathbf{E}\big(\mu \mid \theta(t), \mu(t), s_t, v_t\big). \tag{8}$$

Such an equilibrium exists if the posterior expectation of v_t given such a decision rule is consistent with the equilibrium conjecture of $\hat{v}(a_t) = \alpha_t a_t + \beta_t$. Given the distributional assumptions, one can write this posterior expectation as follows,

$$\mathbf{E}\Big(v_t \mid a_t(1+\kappa\alpha_t^2)\Big) := a_t(1+\kappa\alpha_t^2) \cdot \tilde{s}(\alpha_t, \kappa, \theta(t), \mu(t)) + \iota(\alpha_t, \kappa, \beta_t, \theta(t), \mu(t)), \tag{9}$$

where $\tilde{s}(\cdot)$ will be thought of as determining the sensitivity of the decision rule to v_t and $\iota(\cdot)$ as determining the intercept. Thus the necessary and sufficient condition for a signaling equilibrium to involve revelation is whether there exist an α_t and β_t which satisfy,

$$(1+\kappa\alpha_t^2)\tilde{s}(\alpha_t, \kappa, \theta(t), \mu(t)) = \alpha_t \tag{10}$$

$$\iota(\alpha_t, \kappa, \beta_t, \theta(t), \mu(t)) = \beta_t. \tag{11}$$

These equations resemble those in Equation (3) where, in that benchmark, $\tilde{s}(\cdot) = 1$ and $\iota(\cdot) = \kappa\alpha_t\beta_t$. Similar to that benchmark, whenever there exists a solution to Equation (10), there will exist a unique solution to Equation (11), thus shifting the focus to Equation (10). Further, Equation (10) is independent of β_t. Finally, the sensitivity, $\tilde{s}(\cdot)$, is independent of the means of the beliefs, $\theta(t), \mu(t)$. This independence arises because the sensitivity captures how variation in the decision corresponds to variation in a player's type, which is independent of the mean beliefs in a linear equilibrium. The following lemma formalizes this intuition and provides additional properties of the learning process.

Lemma 1 *Fix κ, $\{\gamma_t\}$, and initial beliefs about θ and μ. Social learning about fundamentals occurs if and only if social learning about preferences occurs. Further, whether or not social learning about fundamentals occurs (symmetrically, preferences) is independent of the realizations of a_t.*

The intuition behind the first statement in the lemma is that if the players socially learn θ, the players must have observed infinitely many periods of informative decisions. Given the knowledge of what θ is, the players can use these infinitely many periods to infer μ and vise versa. Since the conditions for social learning about preferences and fundamentals are identical, for brevity, I will refer to social learning as when the players socially learn both the preferences and fundamentals. Further, the second statement formalizes the intuition that whether an equilibrium involves revelation is determined only by the sensitivity in the conjectured type as a function of the decision. Importantly, that the sensitivity is influenced by the precision of the beliefs in a given period is exactly why the *effective conformity concerns*, corresponding to the endogenous reputational penalty of adaptation, change over time. Adaptation imposes a change in one's perceived type determined by the sensitivity, and the conformity loss is equal to κ multiplied by this sensitivity. To build intuition, if the beliefs about θ are sufficiently imprecise, each player will put comparatively more weight on his signal s_t. As the decision rule now puts a larger weight on s_t than v_t, the sensitivity will be lower because variation in a_t will be ascribed to a player's signal as opposed to his type. In contrast, if the players are sufficiently certain about θ and μ, then each player effectively disregards his signal and his type when computing the posterior expectations of θ and μ, implying that the right-hand side of Equation (8) is approximately equal to v_t and that the sensitivity is approximately equal to 1. One can use this intuition to generate the following lemma, providing a sufficient condition for a failure in asymptotic learning.

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

This lemma states that when the conformity concerns exceed the threshold for revelation in the common knowledge environment, the players are unable to socially learn θ or μ. To gain intuition, note that by Lemma 1 an equilibrium with revelation exists if and only if there exists an α which solves Equation (10). That $\kappa > \kappa^{c.k.} = 1/4$ implies that if the sensitivity were equal to one (or within an ϵ window of 1), then there would exist no solution, as the left-hand side of Equation (10) would be strictly greater than the right-hand side for any α. Hence, for a signaling equilibrium with revelation to exist, the sensitivity cannot converge to one. Further, for social learning to occur, there must exist infinitely many periods of revelation even as the beliefs converge to the truth. However, if the beliefs converge to the truth, the right-hand side of Equation (8) converges to v_t, implying that the sensitivity converges to 1, yielding a contradiction. As a result, when $\kappa > \kappa^{\text{c.k.}}$ the beliefs do not converge and the players face three inefficiencies in the limit. First, as the players use a pooling decision rule in the limit, the players are unable to adapt to their private types v_t. The subsequent two inefficiencies stem from the players utilizing a pooling decision rule based on inaccurate perceptions of *both* θ and μ.

Given Lemma 2, it suffices to analyze $\kappa < \kappa^{\text{c.k.}}$. First note that for any $\kappa < \kappa^{\text{c.k.}}$, there exist initial beliefs such that social learning succeeds. To see why, note that for sufficiently precise beliefs about θ and μ, the noise stemming from adaptation to changes in the perceptions of θ or μ is sufficiently small. As a result, for any ϵ, there exist sufficiently precise beliefs about θ and μ, such that the sensitivity in the decision to v_t is less than $1+\epsilon$. By Equation (10), an equilibrium with revelation exists if and only if $1 + \kappa\alpha_t^2$ multiplied by the sensitivity is equal to α_t. Since $\kappa < \kappa^{\text{c.k.}}$, one can increase the left-hand side by ϵ (corresponding to an increase in κ) and there will still exist a solution. Hence, for sufficiently precise beliefs, there will exist an equilibrium with revelation. Further, as the beliefs in the subsequent period will be more precise, the next period will involve revelation, and by induction, all subsequent periods. Finally, the assumed time fixed-effects in γ_t imply that the players can separately identify θ and μ.[17] This intuition implies that if $\kappa < \kappa^{\text{c.k.}}$, there exist sufficiently precise beliefs for which social learning will succeed. The following lemma formalizes this intuition.

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

The question now turns to whether $\kappa < \kappa^{\text{c.k.}}$ is a sufficient condition for social learning for all possible initial beliefs. Intuitively, for social learning to fail, there must exist a period in which the signaling equilibrium is pooling. That the signaling equilibrium is pooling is equivalent to there not being a solution to $1 + \kappa\alpha_t^2$ multiplied by the sensitivity in the decision rule to v_t equalling α_t (at a high level, the effective conformity concerns

[17]As discussed in Footnote 12, without this assumption the beliefs may converge to an unstable learning outcome.

being larger than the conformity concerns). As one example of why this may occur, note that given a sequence of high decisions each player is unsure if the decisions were high due to a high fundamental state or high average preferences. If the player has a low type, he updates that the fundamental state must be comparatively high. If this inference is sufficiently strong, the sensitivity in the posterior expectation of v_t given a change in the decision may be greater than one despite the noise stemming from s_t suggesting that the sensitivity may be greater than one. When this sensitivity is strictly greater than one, the condition for the existence of an equilibrium with revelation is strictly tighter than $\kappa < \kappa^{\text{c.k.}}$. As a result, there may exist an open set of initial beliefs such that social learning about both preferences and fundamentals fails, despite $\kappa < \kappa^{\text{c.k.}}$. The following proposition formalizes this logic and unifies the previous lemmas.

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

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

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

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

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

This section shows that the condition for social learning-about either fundamentals or the preferences of others-depends on an intuitive fundamental: the magnitude of conformity concerns.[18] I discuss this finding in the context of the empirical literature in Section 4.3.

[18]At this level of generality, the learning outcomes are not monotone in κ. For instance, a

4.2 Comparative Statics

In the previous subsection the players possess uncertainty about both a fundamental state and the preferences of others. Allowing for both types of uncertainties allows for the comparison between interventions addressing misperceptions of the preferences of others and misperceptions about the fundamental state in Section 5. However, the finding that social learning about fundamentals occurs if and only if conformity concerns are sufficiently small continues to hold absent uncertainty about the preferences in the population. Further, the analysis absent such uncertainty allows for additional comparative statics about the long-run behavior of the players.

In this subsection, I assume, μ, the average preferences of the community is common knowledge and without loss of generality equal to zero. Further, as the learning problem has only one dimension of uncertainty, one can normalize $\gamma_t = 1$ without loss of insights (see Footnote 12). As such, a sufficient statistic for the history is the current public belief about θ, $\theta(t) \sim N(\bar{\theta}(t), \tau_{\theta,t})$.

The evolution of $\theta(t)$ uniquely determines the equilibrium behavior of the players. Further, as argued in Section 4.1, the mean of $\theta(t)$, $\bar{\theta}(t)$, does not influence whether or not an equilibrium will be revealing nor the degree of revelation. Thus, the equilibrium dynamics are determined by the precision of $\theta(t)$, $\tau_{\theta,t}$. In this simplified model, the effect of greater uncertainty about θ is that deviations in the decision are increasingly ascribed to s_t as opposed to v_t. As a result, the greater the uncertainty, the lesser the effective conformity concerns. These intuitions combine to generate the following proposition.

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

This proposition shows that conformity concerns impacts not only the binary outcome of social learning, but also the degree of asymptotic learning. The intuition behind this result is that uncertainty about θ, in this simplified environment, always decreases the effective conformity concerns. As a result, if the equilibrium involves revelation in the environment with common knowledge, then the equilibrium is revealing in all periods with uncertainty about θ. This implies that if $\kappa \leq \kappa^{c.k.}$, the players learn θ for any initial beliefs. In contrast,

marginally higher value of κ alters whether the decision rule places marginally more weight on s_t or v_t, which results in differential beliefs in the subsequent period. In either analysis with only one dimension of uncertainty, the learning outcomes are strictly monotone with respect to κ.

if $\kappa > \kappa^{c.k.}$, then in the common knowledge environment the players cannot adapt to their private information. As social learning necessitates adaptation even as the beliefs become arbitrarily precise, the players necessarily stop adapting to their private information before learning occurs.

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

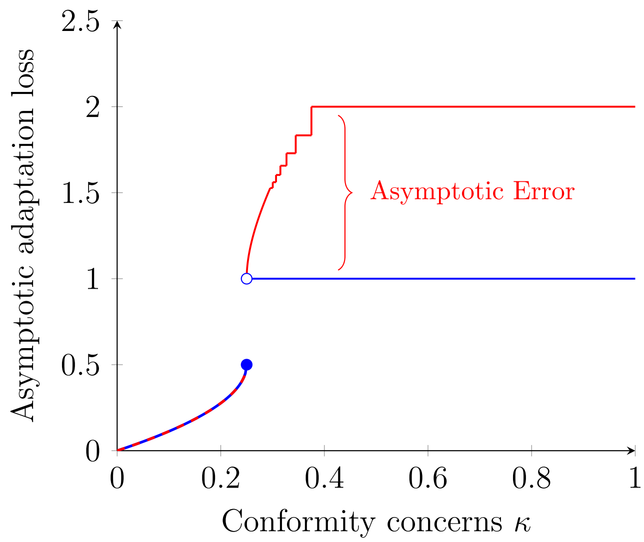

Figure 2: Asymptotic Adaptation Loss

In Blue is the asymptotic adaptation loss in the benchmark. In Red is the asymptotic adaptation loss when the players have common knowledge about μ, the average preferences of their peers, but have prior $\theta \sim N(0,1)$ about the fundamental state. These two functions coincide when $\kappa \leq \kappa^{\text{c.k.}}$ but differ for higher values of conformity concerns. The difference between these two functions is due to the players pooling on $\bar{\theta}(t)$, which will not equal θ, when the players fail to socially learn θ.

This simplified model showcases how conformity concerns impede efficient learning. Despite the players having access to continuous decisions and sufficiently informative signals, the players fail to learn the fundamental state for any prior beliefs when the conformity concerns are sufficiently high. Further, in addition to the extensive margin of whether the beliefs perfectly converge to the truth, the conformity concerns effect the intensive margin of the precision of the beliefs: the greater the conformity concerns, the more imprecise beliefs the players ultimate harbor.

4.3 Applied Relevance

My model has predictions for both the asymptotic efficiency of decisions and whether the players will learn the average preferences of their peers. I now argue that my predictions are more consistent with the empirical literature than the existing theoretical literature discussed in Section 1.1.

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

Continuous Decisions: As discussed in Lee (1993), examples of successful social learning include financial markets and business forecasting where investment decisions are continuous. The literature notes that despite a potential for initial erroneous mistakes and herding, that firms asymptotically firms learn whether a given asset is valuable. In contrast, my model predicts that the conformity concerns must be low for efficient decisions asymptotically. Returning to the case study of financial markets, one might think the conformity concerns are low relative to the financial stakes. Given these low conformity concerns, my model produces a similar prediction to the historical literature. However, if decisions are continuous, but the conformity concerns are high, such as alcohol consumption (cf. Prentice and Miller, 1993), drug use (cf. West and O'Neal, 2004), and many others, then in line with my model, we see more inefficiencies and worse beliefs amongst the community.

Unboundedly Informative Signals: The social learning literature defines a signal as unboundedly informative signals if with positive probability the player is arbitrarily certain of the optimal decision after observing one's private signal. If these signals occur, then even if the players are herding on a wrong decision, when such a signal occurs, the player who received such a signal will break the herd and choose the correct decision. This observation is in contrast to the famous conformity experiment in Asch (1953). Participants were grouped and shown a series of lines, then asked to identify the one matching a reference line. Unbeknownst to the participants, the experimenters planted an actor into the group to deliberately provide incorrect answers. Without the actors, success exceeded ninety-nine percent, but with the actors, over seventy-five percent of participants conformed. The social learning literature predicts that individuals should not copy the actor because each individual can identify the correct answer. In contrast, my model predicts that if conformity concerns are large, the individuals will copy the actor's incorrect decision.[19] Finally, Franzen and Mader (2023) replicated the original study and found that monetary incentives decreases the probability of conformity by 13 percentage points. Consistent with my model, these incentives increase the importance of adaptation, resulting in less conformity.

Pluralistic Ignorance: My model predicts that "pluralistic ignorance" can arise in equilibrium. Pluralistic ignorance is defined as, "a situation in which group members systematically misestimate their peers' attitudes" (Miller, 2023). In a review article, Bursztyn and Yang (2022) document that such misperceptions lead to inefficient social norms and are rampant, occurring in a wide range of environments: political movements, macroeconomic expectations, vaccination status, and many others.

The extent of pluralistic ignorance corresponds to the magnitude by which individuals systematically misestimate the preferences of their peers, for a given realization of their peers true preferences.[20] In the model, if the public beliefs about μ converge to the truth, then (tautologically) every player correctly predicts the preferences of their peers. In contrast, if the public beliefs do not converge, then there exists uncertainty about μ, implying that each player's estimate of his peers' preferences combines both the public beliefs about μ and his private preferences, v_t, which are predictive about μ. In such cases, the distribution of estimates will be non-degenerate, and with a probability equal to one, will not be perfectly centered around μ. Further, the greater the uncertainty in the public beliefs, the more likely such beliefs are centered around the prior beliefs about μ as opposed to the true value. Thus, one can view the uncertainty about the public beliefs about μ (the inverse of $\tau_{\mu,t}$) as describing the expected degree of pluralistic ignorance.

While there exist numerous behavioral explanations for pluralistic ignorance, the model presented in this paper provides an additional explanation: the desire for conformity necessitates "self-censorship in public discourse" (cf. Loury, 1994), resulting in insufficient

[19] In my model, decisions are continuous. However, one could alter my model to consider discrete decisions, and indeed one would find that players would conform even in the presence of unboundedly informative signals.

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

information for others to gauge the views of the public.[21]

5 Policy Interventions

In this section, I first review two case studies about interventions. Next, I extend the model to analyze informational interventions and show that when conformity concerns are high, interventions addressing misperceptions of the average preferences in the community outperform interventions addressing misperceptions of the fundamental state. Finally, motivated by the literature in social psychology on pluralistic ignorance, I extend the model to consider changing preferences and connect this finding to peer-oriented interventions.

5.1 Empirical Examples

In this subsection I detail two empirical studies about interventions from social psychology and economics, respectively.

Case 1: Alcohol Use on Campus: Prentice and Miller (1993) conducted a survey amongst Princeton undergraduates to show that students over-estimate their peers' preferences towards alcohol by 32 percent and that such misperceptions are correlated to the over-consumption of alcohol on campus.[22] Given the results in Prentice and Miller (1993), Schroeder and Prentice (1998) causally tested whether these misperceptions are the primary cause of excessive drinking on campus as opposed to potential misperceptions of the deleterious health consequences of excess drinking. To do so, the authors divided incoming college students into two groups. The first group received information about the health effects of alcohol consumption (θ in my model), and the second group received information about their peers' preferences towards alcohol (μ in my model). At the end of the semester, students were surveyed about their drinking behavior, and the intervention targeting misperceptions of peers' preferences reported 40 percent less drinking than the intervention addressing health consequences.

[21] This theory finds empirical backing in Braghieri (2021) which documents that participants are likely to skew their answers to politically sensitive questions in the direction of public support when these answers are public and that such skewing decreases the information from any given statement. In contrast to Braghieri (2021), I allow for uncertainty about the direction of the public support. For instance, college students may be uncertain how "cool" drinking is. Allowing for such uncertainty allows for a tighter connection with the case studies regarding pluralistic ignorance.

[22] Prentice and Miller (1993) even notes, "Princeton reunions boast the second highest level of alcohol consumption for any event in the country after the Indianapolis 500," implying that such perceptions are sufficiently strong to continue a decade after graduation at a reunion and have deleterious outcomes for the community.

Case 2: Women Working Outside the Home: A recent randomized experiment from Bursztyn et al. (2020a) find that, in the context of Saudi Arabia, the vast majority of young married men privately support women working outside the home (WWOH) and substantially underestimate support by other similar men. Further, the fact that the city has low levels of WWOH, suggests that such misperceptions may be impeding efficient behavior. The authors then randomly assign information about the correct perceptions of one's peers to the participants and show that correcting the misperceptions increases men's willingness to have their wives work by 36 percent. In a final step towards showing that misperceptions of peers' preferences are the main inhibitor of efficient decision-making, Bursztyn et al. (2020a) consider whether misperceptions about the fundamental state may be the root cause, noting, "if so many people, in fact, support WWOH then there are probably many firms willing to hire women for jobs outside the home" (p. 3018). They show that information about preferences of their peers caused no updated inference in the number of jobs for women, consistent with the fact that the effect is not caused through updated beliefs about the fundamental state.[23]

5.2 Modeling Interventions

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

Before analyzing "common-knowledge" interventions, I analyze "private interventions." One can think of a private intervention as giving player t access to additional information; however, such information is private and is not accounted for by the community when inferring player t's preferences given his decision. Without formally stating the result, one can see that such an intervention has no ability to break a pooling equilibrium nor influence which decision the players pool on. To see why, suppose player t is told the value of the fundamental state, θ. Given that each player wants to match his decision to the fundamental state, player t has an identical incentive to adapt to θ as a hypothetical player who received a signal whose implied posterior mean of θ matches the fundamental state. Further, since the signal distribution has full support, and the hypothetical player cannot adapt to such

[23] Bursztyn et al. (2020a) include a model with a few key differences to mine. First, in Bursztyn et al. (2020a) agents are endowed with incorrect beliefs about their peers' preferences. In contrast, these misperceptions arise in equilibrium in my model. Second, in Bursztyn et al. (2020a) decisions are discrete, whereas my paper shows learning may fail for continuous decisions. Finally, my model allows for misperceptions about θ and μ and gives conditions when information about μ is preferred. In contrast, in Bursztyn et al. (2020a) θ (which can be viewed as the economic benefits of WWOH) is common knowledge and the only possible interventions are about μ.

information, neither can player t. Thus, the equilibrium in period t remains identical. Finally, as such information was private information to player t, and no change in behavior occurs in period t, then no change in behavior follows for any subsequent periods.[24]

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

5.3 The Effects of Interventions

I begin with a definition of when a signaling equilibrium with pooling is fragile. I call a pooling decision rule "fragile" to an individual-oriented intervention with n pieces of information if there exists a hypothetical public disclosure of n i.i.d. signals about θ that are distributed identically to s_t which cause the equilibrium in period t to be non-pooling when it would otherwise be pooling. Similarly, it is fragile to a peer-oriented intervention with n pieces of information if n i.i.d. signals to v_t causes an equilibrium to be non-pooling when it would otherwise be pooling. This definition mirrors the definition of fragility in Bikhchandani et al. (2021) but is augmented to allow for a signal about μ.

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

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

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

[24]In support of this theory, Tevyaw et al. (2007) shows that the magnitude of the reduction in alcohol use was 3 times larger on average for group-level interventions than individual-level interventions.

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

The intuition behind the first result is that if $\kappa < \kappa^{c.k.}$, then when there is complete information the signaling equilibrium involves revelation. Further, that the equilibrium is pooling despite a $\kappa < \kappa^{c.k.}$ is necessitated by a strong negative correlation between the beliefs about θ and μ. When this correlation is strong, each player negatively updates about θ given his type, v_t. As a result, when the correlation is negative, changes in v_t cause a smaller change in a_t, increasing the reputational penalty from adapting to one's private information (as this is increasingly ascribed to v_t). A sufficiently large amount of information about either θ or μ will weaken the negative correlation in the beliefs about θ and μ, allowing the players to adapt to private information once more.

The second result says that if the conformity concerns are high, then giving information about only one dimension of uncertainty will be unsuccessful in spurring social learning. The proof for this result is precisely Proposition 2, which states that, for any prior beliefs about θ and μ, the players beliefs about θ and μ cannot converge to the truth. The implication of this result is that even if a social planner could design the perfect individual-oriented intervention, the players will necessarily continue to pool on inaccurate perceptions of their peers preferences.

Finally, the intuition for the final result comes from the different effects of these interventions on the effective conformity concerns. If there is no correlation between the beliefs about θ and μ the first-order condition defining a player's decision in Equation (8) simplifies to:

$$a_t(1+\kappa\alpha^2) = \gamma_t \mathbf{E}(\theta|\theta(t), s_t) + v_t + \kappa\alpha\mathbf{E}(\mu|\mu(t), v_t). \tag{12}$$

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

In contrast, a peer-oriented intervention has two effects. First every players wants to be perceived as the average type. Consequently, when $\tau_{\mu,t}$ is low, players with different preferences will have different perceptions of what the average type is. Intuitively, if each player has different perceptions of the population's average, then each player will adapt his decision to his private preferences because doing so will ensure the public perception of his preferences will be in line with the population's average. This logic implies that when $\tau_{\mu,t}$ is low, the players have an added incentive to adapt.

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

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

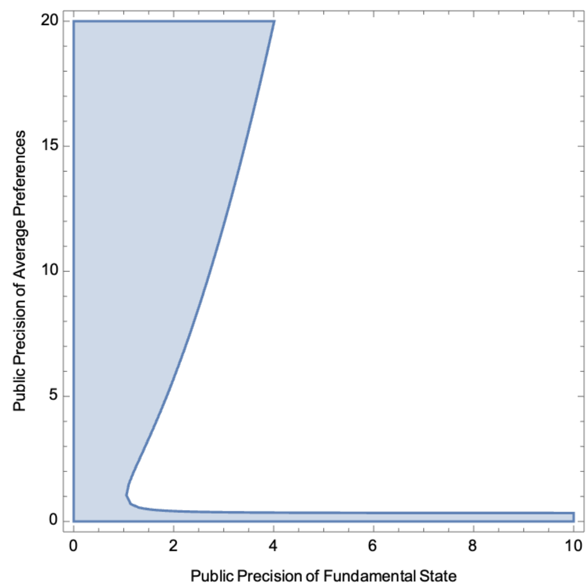

Figure 3: When the Signaling Equilibrium Involves Revelation

In the figure, the x-axis corresponds to $\tau_{\theta,t}$, the public precision of the fundamental state, and the y-axis corresponds to $\tau_{\mu,t}$, the public precision of the preferences of others. The shaded region corresponds to when the signaling equilibrium involves revelation. The non-shaded region corresponds to values where the signaling equilibrium is pooling. In the figure $\kappa = \gamma_t = 1$, which correspond to the weight of conformity and the relative value players place on adapting to the fundamental state.

However, even if neither a peer-oriented intervention nor an individual-oriented intervention break a pooling equilibrium, these interventions will influence which decision the players pool on. To gain intuition into the forces behind the change, I consider the following situation: suppose the equilibrium in period 1 was revealing and thereafter the decision rule 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^*, where for simplicity I assume $\gamma_2 = 1$ to derive:

$$a^*(a_1) = \mathbf{E}(\theta|a_1) + \mathbf{E}(\mu|a_1). \tag{13}$$

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

$$a_1 = \lambda_\theta s_1 + \lambda_v v_1, \tag{14}$$

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

Equation (13) for all subsequent periods.

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

$$\Delta(\mu) = \mu\left(1 - \frac{\lambda_\theta\lambda_v}{\frac{1+2\tau_\theta}{1+\tau_\theta}\lambda_\theta^2 + \frac{\tau_\theta}{1+\tau_\mu}\lambda_v^2}\right) + \alpha_\mu a_1 \tag{15}$$

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

for some constants α_θ, α_v*.*

To understand the expressions above, let us now consider the effect of an individual-oriented intervention revealing θ, where a symmetric analysis occurs for μ. Upon revealing θ, the updated equilibrium perception of θ is θ. Further, the players re-evaluate the perception of μ as a function of both θ (the second term in the parentheses of Equation (16)) and a_1. The object of interest is how much the players decision changes with respect to μ. One can see that given information that θ is positive (respectively, negative) the players update that μ is negative (respectively, positive). Further, this update could be larger in magnitude than the update about the value of θ. Specifically, these cases occur when τ_μ is small (i.e., the players are uncertain about their peers' true preferences). Such a counter-update provides one rationale why the individual-oriented interventions have a small (if not negative) effect on behavior.

5.4 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 and the fundamental state. Proposition 4 suggests that peer-oriented interventions may be preferred due to their ability to break a pooling equilibrium. Further, Proposition 5 suggests that even a perfect individual-oriented intervention alone may fail to shift the pooling decision in the direction of efficiency.

These predictions are broadly consistent with the results in Schroeder and Prentice (1998) and Bursztyn et al. (2020a) for two reasons. First, interventions addressing the misperceptions of the preferences of others are preferred. Second, in both settings conformity concerns are arguably high. If instead conformity concerns were low (or in the limit equal to zero), then similar to the theoretical literature, the optimal intervention would be individual-oriented. When the conformity concerns are low, the preferences of one's peers are less relevant and individual-oriented interventions allow the players to reach an efficient decision faster.

5.5 Changing Preferences and Peer-Oriented Interventions

The social psychology literature notes that "a society's perception of itself tends to lag behind actual changes in people's private beliefs and values," and argues this lag necessitates peer-oriented interventions to improve decision-making (Miller, 2023). To address these phenomena, I generalize the main model to allow μ to be time dependent and follow an autoregressive process parameterized as,

$$\mu^*(t) = \rho\mu^*(t-1) + (1-\rho)\psi_t, \tag{17}$$

where ψ_t is independent across time and distributed as $N(0, \tau_\mu)$, with τ_μ being the same τ_μ as in Section 2 where $\mu \sim N(0, \tau_\mu)$. Making this equivalence allows the ex-ante uncertainty about $\mu^*(t)$ to be equal to that of μ for all t. When the average preferences change over time, one can analyze how decisions respond to preference changes. The first implication of changing preferences is that switches to pooling decision rules are never permanent.

In the primary analysis, once the players move to a pooling decision rule, the players pool for all subsequent periods. However, given the changing population average, if the players pool in all subsequent periods, the public beliefs eventually converge back to beliefs where revelation occur. Finally, in this extension, the players utilize the decision rule with revelation for arbitrarily many periods and θ remains fixed, implying that it is without loss to assume that θ is common knowledge and fixed at zero when analyzing the asymptotic behavior of the community. The following proposition analyzes the asymptotic behavior of the players as a function of the conformity concerns.

Proposition 6 (**Shifting Preferences**)
Suppose μ follows an autoregressive process as defined in Equation (17) and denote by $\tau_{\mu,t}$ the precision about the public beliefs about $v^*(t)$ in period t. The signaling equilibrium in period t involves revelation if and only if

$$\tau_{\mu,t} < \frac{\kappa^{c.k.}}{\kappa - \kappa^{c.k.}}. \tag{18}$$

As a result, there exists a threshold $\kappa^* \in \left(\kappa^{c.k.}, \kappa^{c.k.} \cdot (\tau_\mu + 1)/\tau_\mu\right)$ such that

1. *If $\kappa \leq \kappa^*$ the signalling equilibrium involves revelation in all periods and $\tau_{\mu,t} \to \tau(\rho)$. Further, $\tau'(\rho) < 0$ and $\lim_{\rho \to 0} \tau(\rho) = \infty$.*

2. *If $\kappa \in \left(\kappa^*, \kappa^{c.k.}(\tau_\mu + 1)/\tau_\mu\right)$ the signalling equilibrium involves revelation (i.e., the precision in the beliefs is less than the right-hand side of Equation (18)) for infinitely many periods and the players pool for infinitely many periods (i.e., the precision in the beliefs is greater than the right-hand side of Equation (18)), and $\tau_{\mu,t}$ does not converge.*

3. *If $\kappa \geq (\tau_\mu + 1)/\tau_\mu$, the signalling equilibrium is pooling in all periods and $\tau_{\mu,t} = \tau_\mu$ in all periods.*

Before analyzing the asymptotic results, let us build intuition for when the signaling equilibrium in period t involves revelation. Note that in any equilibrium with revelation a player's type is fully revealed in equilibrium, thus $\tau_{\mu,t}$ does not impact the posterior beliefs about a player's type. As a result, the only effect of an increase in $\tau_{\mu,t}$ is that each player's beliefs about μ places less weight on his own type v_t. As each player wants his perceived type to equal μ, and thus chooses a decision which is responsive towards his beliefs about μ, an increase in $\tau_{\mu,t}$ decreases each player's incentive to respond to his private type v_t. Such a force is seen in Equation (18), which notes that as $\tau_{\mu,t}$ increases the condition on κ for an equilibrium with revelation to exist becomes more stringent.

Given Equation (18), it is immediate that if $\kappa \leq \kappa^{\text{c.k.}}$, then the signaling equilibrium would involve revelation for all periods. If the signaling equilibrium involved revelation for all periods, $\tau_{\mu,t}$ would monotonically increase to a constant $\tau(\rho) < \infty$, implying the players have imprecise beliefs about $v^*(t)$. Note that the players can never perfectly infer $v^*(t)$ because the average preferences change in each period. Thus the condition for an equilibrium with revelation in all periods is determined by κ^*, which binds Equation (18) when $\tau_{\mu,t} = \tau(\rho)$.

When the conformity concerns are greater than this value but small enough such that the equilibrium involves revelation in period 1, the players oscillate between a decision rule with revelation and pooling in the limit. This occurs because as more periods with revelation occur (respectively, pooling), the beliefs become more precise (respectively, imprecise), eventually necessitating a pooling (respectively, revealing) equilibrium. As a result, the beliefs do not converge, but rather oscillate around the value $\tau_{\mu,t}$ which binds equation (18). This value is decreasing in κ, implying that higher values of the conformity concerns implies worse beliefs in the limit, and these worse beliefs are caused by a greater number of periods utilizing a pooling equilibrium.[25] Finally, if the conformity concerns are large enough that

[25]A precise characterization of the asymptotic distribution of $\tau_{\mu,t}$ when $\kappa > \kappa^*$ is challenging

the players pool in period 1, then, by induction, the players pool in all periods.

What this proposition implies is that, all else equal, groups who have stronger weights on conformity concerns wait longer to adapt to the underlying conditions. This result is broadly consistent with the social psychology literature on pluralistic ignorance and how norms change. In a review article, Miller (2023) states, "widespread changes in private attitudes change are not sufficient for social norm change. The group's recognition that its collective attitudes have changed is also necessary. Without this recognition, norm change will be impeded." One can interpret this through the model as follows: if the players are pooling in period t on a low decision due to a low belief $\mu^*(t)$, despite a high value for v_{t+1} signaling a change in private attitudes has occurred, this change is *not* sufficient to change the decision rule. Rather, enough periods must pass for the group to be certain that their collective attitudes have changed.

Such a process where the players utilize a responsive decision rule for some number of periods before switching to a pooling decision rule and vice versa, is also in line with the psychology literature on changing norms. For instance, Miller and Prentice (1994) suggest that one major source of pluralistic ignorance is a "conservative lag" whereby opinions change but not decisions. This can be viewed as the periods in which the players utilize a pooling decision rule despite their values changing. As the state has likely changed over this time frame, the subsequent player's decision may differ from what the previous pooling decision was. Such a change exemplifies the "liberal leap" also described in Miller and Prentice (1994).[26] This extension shows how groups that have higher conformity concerns will have greater degrees of pluralistic ignorance and adapt slower to changes in private attitudes. Finally, note that such changing preferences are an additional motivation behind the peer-oriented interventions considered in Section 5.2. Without these interventions, the social norms will lag behind the true attitudes in the population, causing inefficiencies.

6 Extensions

In this section I consider three extensions that serve as robustness checks for the assumptions in the main analysis. First, a common assumption in the social learning literature is that each player is short-lived. While long-lived players traditionally are able to observe more signals, and thus make more efficient decisions absent conformity concerns, the

because $\tau_{\mu,t}$ evolves according to a non-continuous discrete dynamical system.

[26]The economics literature produces similar findings. For instance, an immediate implication of the model is that when the autoregressive process is more volatile, the players will spend less periods in the pooling decision rule. This result is in line with Giuliano and Nunn (2021), which shows that populations in uncertain climates have less persistent norms.

conformity concerns are amplified through a ratchet effect. Second, I show that when off-path beliefs satisfy D1 from Cho and Kreps (1987), non-linear decision rules fail to exist in the benchmark environment in Section 3. Finally, I discuss how the results of the model generalize to alternative distributional assumptions with non-linear decision rules.

6.1 Long-Lived Players

This subsection analyzes the incentives of a longer-lived player. To model such a phenomena, I continue to index time by $t \in \{0, 1, \ldots\}$, but label the players by $i \in \{0, 1, \ldots\}$. I denote by $t(i)$ the first period in which player i appears. Further, each player, i, continues onto the next period with uniform probability p and with probability $1 - p$ is replaced by player $i + 1$. Each player has a discount factor $\delta < p$, which includes both the probability of continuation and inter-temporal discounting. The model in the primary analysis considers $p = 0$, and as a result $\delta = 0$, and each player makes a single decision. When the players are long-lived, the utility for player i is as follows:

$$u^{t(i)} + \delta u^{t(i)+1} + \delta^2 u^{t(i)+2} + \ldots. \tag{19}$$

This game featuring persistent private information entails well-known non-trivial modeling choices. Assuming θ is common knowledge greatly simplifies the analysis. Further, once θ is assumed to be common knowledge, it is without loss of generality to consider $\theta = 0$.

The final simplification I make is a restriction to the following class of equilibria that mimic the characterization in the primary analysis. In period $t(i)$, player i utilizes a linear decision rule. If there exists a linear decision rule with revelation, such a decision rule is utilized in period $t(i)$. As player i's preferences are then fully revealed on-path, in all subsequent periods player i utilizes a pooling decision rule as determined by the posterior mean of player i's perceived preferences. On path, this corresponds to player i's true preferences and is hence socially optimal. When there does not exist a revealing linear decision rule that constitutes an equilibrium in period $t(i)$, the decision rule is fully pooling in period $t(i)$. As no new information is learned, period $t(i)+1$ is equivalent to $t(i)$, and thus a pooling decision rule will be used in period $t(i) + 1$, and by induction, for all subsequent periods for player i. Note that this set of equilibrium refinements prescribes a unique decision rule for each player because the pooling decision rule is always an equilibrium.

Given the equilibrium selection, the trade-off for player i in period $t(i)$ is that a higher decision today allows for a higher decision to be taken in all subsequent periods because a higher decision increases the perception of player i's preferences. However, the perception being higher also implies a higher conformity loss in all subsequent periods. The following proposition states how this trade-off changes as the discount factor, δ, increases.

Proposition 7 (**Long-Lived Players**)
When players are long-lived with a discount factor, δ, have a conformity weight, κ, and θ is common knowledge, then an equilibrium with revelation exists if and only if:

$$\kappa < \kappa^{c.k.}(1-\delta)\frac{\tau_{\mu,t}+1}{\tau_{\mu,t}}, \tag{20}$$

where $\tau_{\mu,t}$ denotes the precision about the public beliefs about the average preferences.

To understand this expression, first note that when $\delta = 0$ the condition for an equilibrium is equivalent to the condition in Proposition 6. Let us now turn to how the discount factor, δ, affects the condition. To understand why an increase in δ increases the effective conformity concerns, note that, in equilibrium, choosing a marginally more conforming decision in period $t(i)$ has only a second-order effect on the adaptation loss in all subsequent periods. Recall, in equilibrium, player i's preferences are correctly inferred, and thus the decision in all future periods is v_i. If player i selected a marginally more conforming decision, this deviation would generate only a second-order loss in adaptation in subsequent periods, as the adaptation loss is quadratic. However, a marginally more conforming decision in period $t(i)$ has a first-order impact on the conformity loss in all subsequent periods. Thus, an increase in the discount factor effectively scales the conformity concerns, as there is no future benefit to adaptation; however, there is a greater future benefit from conformity.

Proposition 7 generates the stark prediction that the degree of misperceptions about others' preferences is shaped by the players' discount factor. Ignoring integer constraints, the players cease utilizing decision rules with revelation when $\tau_{\mu,t}$ binds Equation (20). Solving Equation (20) implies that when the players' discount factor is higher, the players learn less about the preferences of others. This result adds a competing force to those suggested by the social learning literature. In that literature, long-lived players have an ability to observe more data and thus make more accurate decisions. This extension highlights that when conformity concerns are present, long-lived players might make worse decisions.

6.2 Non-Linear Equilibria

The main analysis analyzed signalling equilibria: linear and socially optimal Perfect Bayesian Equilibria. This subsection shows that while non-linear Perfect Bayesian Equilibria may exist, they do not satisfy the D1 refinement from Cho and Kreps (1987).[27] This

[27] Cho and Kreps (1987) define D1 for signalling games. Their definition applies to my setting if one views the reputation as stemming from the player interacting with one randomly drawn member

refinement states that off-path beliefs are concentrated on the types who have the largest incentive to deviate to such a decision. As in Section 3, when θ and μ are common knowledge it suffices to consider a static version of the game and drop any time-dependence. Further, Equation (1) reduces to:

$$u(v,a) = -(a-v)^2 - \kappa \int b^2 \phi(b,a) db. \tag{21}$$

In doing so, I first restate the definition of a "central pooling equilibrium" from Bernheim (1994). A central pooling equilibrium is an equilibrium in which $a(v) = c \quad \forall v \in [\underline{v}, \bar{v}]$ where $\underline{v} \leq 0 \leq \bar{v}$ and $a(v)$ is strictly monotone when $v \notin [\underline{v}, \bar{v}]$. The lemma below shows that any equilibrium which satisfies D1 is a central pooling equilibrium.

Lemma 4 (**Class of Equilibria**)
Any equilibrium satisfying D1 is a central pooling equilibrium. In this equilibrium, $a(v) = c^* \quad \forall v \in [\underline{v}, \bar{v}]$ *where* $\underline{v} \leq 0 \leq \bar{v}$*. Further, for* $v \notin [\underline{v}, \bar{v}]$ $a(v)$ *is continuously differentiable with a derivative that satisfies:*

$$a'(v) = \frac{\kappa v}{v - a(v)} > 0. \tag{22}$$

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

The primary difference relative to Bernheim (1994) is that I assume that the support of the distribution of v equals the real line whereas Bernheim (1994) assumes the support of v equals a bounded interval. As such, the solution to Equation (22) must exist over the real line in my setting, but need only exist over a bounded interval in Bernheim (1994).[28] As

of the community who gives the player a reward in accordance with this community members' perception of the player's type.

[28]In Bernheim (1994), the decision rule from this differential equation is non-linear and equilibria with partial revelation exist. The reason for the stark difference is that Bernheim (1994) considers a bounded support of v. If, for example, 1 is the supremum of the support, then the definition of a central pooling equilibrium implies that the player with $v = 1$ has his preferences revealed in equilibrium, and thus might as well choose their preferred decision. Hence, the differential equation in Bernheim (1994) has an initial condition of $d(1) = 1$. In the setting with an infinite support, no player chooses $d(v) = v$ outside the central pool, which generates different conditions for equilibrium existence. This distinction is further discussed in Appendix B.

there are multiple solutions to the differential equation (cf. Figure 1), one cannot rule out non-linear equilibria with an equilibrium uniqueness result. To do so necessitates solving the differential equation in Equation (22) in closed form, which is done in Appendix A to prove the following proposition.

Proposition 8 (**Non-existence of Non-Linear Equilibria**)
Any equilibrium satisfying D1 is linear.

This result gives support for the restriction to linear equilibria in the main analysis. Further, this result implies that if $\kappa > \kappa^{\text{c.k.}}$, no equilibrium with revelation satisfies the D1 refinement.

6.3 General Distributions

Throughout the analysis I focused on the Gaussian distribution which allowed for closed-form solutions and precise comparative statics. In this subsection, I will discuss to what extent these results generalize to different distributions. Recall that the analysis in the previous subsection, which assumed no uncertainty over the fundamental state θ, or the average preferences of the population μ, allowed for any distribution over v_t, the preferences of player t, with a continuous density with support equal to the real line. In that analysis the equilibrium is pooling if and only if the conformity concerns, κ, exceed $\kappa^{\text{c.k.}}$.

The environment with general distributions and both dimensions of uncertainty is intractable.[29] Therefore, I will conduct two separate analyses, each focusing on a different dimension of uncertainty. In this subsection I will provide the intuition for the case where θ is common knowledge and the players are learning the preferences of their peers μ, and Appendix B contains a parallel analysis when θ is uncertain and μ is known. Recall that the analysis when θ is common knowledge with the Gaussian assumptions is contained in Section 5.5 when $\rho = 0$ or Section 6.1 when $\delta = 0$. In both analyses, when players have more uncertainty over their surroundings, the players can coordinate on an equilibrium with revelation for higher values of κ. This can be used to show that a higher κ implies the players switch to a pooling equilibrium with less precise beliefs and, thus, ultimately harbor these less precise beliefs. This subsection analyzes to what extent this point relies on the properties of the Gaussian distribution.

When the distribution of μ and v_t are Gaussian, the public beliefs about μ at time t will satisfy $\mu(t) \sim N(\bar{\mu}(t), \tau_{\mu,t})$, for some mean $\bar{\mu}(t)$ and precision $\tau_{\mu,t}$. Further, the updating

[29] In Section 9.4, I detail the analytical challenges that arise when both factors are present. Intuitively, the differential equations governing the decision rules transform into partial differential equations, significantly complicating the analysis.

rule for the conditional expectation of μ given the realization of v_t has the following closed-form expression as shown in the utility of player t below:

$$-(a_t - v_t)^2 - \kappa\Big(\hat{v}_t(a_t) - \underbrace{\frac{v_t + \bar{\mu}(t)\tau_{\mu,t}}{1+\tau_{\mu,t}}}_{\mathbf{E}(\mu|v_t)}\Big)^2. \tag{23}$$

Here, the constant equal to 1 is the precision of player t's preferences relative to μ. Uncertainty over μ (i.e., a lower value of $\tau_{\mu,t}$) gives player t an additional incentive to respond to v_t, which implies a higher threshold value of κ for the existence of an equilibrium with revelation when $\tau_{\mu,t}$ is higher. This point does not rely on the Gaussian assumption. Let us consider any full support and atomless distribution of v_t. Further, I denote by $g_t(v_t|a_1,\dots,a_{t-1}) = \mathbf{E}(\mu|v_t,a_1,\dots,a_{t-1})$, which, with an abuse of notation, will be denoted as $g_t(\cdot)$. Given this notation, the generalization of the differential equation in Equation (8), which describes the decision rule in any equilibrium with revelation satisfying D1, is:

$$a_t'(v_t) = \frac{\kappa\big(v_t - g_t(v_t)\big)}{v_t - a_t(v_t)}. \tag{24}$$

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

Proposition 9 (**General Distributions with θ known**)
Denote by $i_t := \inf_x g_t'(x) \leq \sup_x g_t'(x) := s_t$, where $g_t(x) = \mathbf{E}(\mu|x,a_1,\dots,a_{t-1})$. There exists an equilibrium with revelation that satisfies D1 if the conformity concerns, κ, satisfy

$$\kappa \leq \frac{\kappa^{c.k.}}{1-i_t}. \tag{25}$$

Further, no equilibrium with revelation satisfies D1 if

$$\kappa > \frac{\kappa^{c.k.}}{1-s_t}. \tag{26}$$

This proposition gives a separate necessary and sufficient condition for the existence of an equilibrium with revelation. In the Gaussian analysis, $g_t(v_t)$ is linear which implies $s_t = i_t$. When $s_t = i_t$, these necessary and sufficient conditions coincide, implying these bounds are tight. The basic insight is that an equilibrium with revelation exists if and only if the differential equation in Equation (24) has a solution. For a solution to exist, $a_t(v_t)$ must never exceed v_t to ensure that the denominator is non-zero. Further, a higher value of conformity concerns, κ, uniformly increases the solution to the differential equation in Equation (24) given any initial condition. For this reason, when κ is sufficiently low (respectively, high) an equilibrium with revelation exists (respectively, does not exist).

This proposition also characterizes the effect of population uncertainty on the degree of revelation. One can view both s_t and i_t as measures of the degree of population uncertainty. Similar to the Gaussian analysis, this greater population uncertainty gives an added incentive for the players to adapt to their private information and ultimately increases the cutoff κ^*.

As a result of this proposition, one can recover similar results to the main analysis. Namely, whenever the beliefs become sufficiently precise relative to the conformity concerns, the players switch to a pooling equilibrium based on imprecise perceptions of their peers. In the Gaussian analysis, $s_t = i_t$ and these values follow a deterministic process. This determinism implied a monotone relation between κ and the asymptotic uncertainty over the preferences of one's peers. In this general analysis, such a claim need not be true.[30] However, the result that pluralistic ignorance exists in all equilibria if and only if κ exceeds $\kappa^{\text{c.k.}}$ continues to hold. To see why, note that if $\kappa < \kappa^{\text{c.k.}}$ the sufficient condition for the existence of an equilibrium always holds, and thus with an infinite number of periods with revelation, the players learn the true average. In contrast, if $\kappa > \kappa^{\text{c.k.}}$, then there must exist sufficiently precise beliefs such that the necessary condition for equilibrium existence fails to hold.

7 Conclusion

This paper studies how conformity concerns impact social learning and what interventions are effective when social learning fails. To do so, I enrich a standard model of social learning by adding: (i) a player's desire to adapt to not only a fundamental state but also his private preferences, (ii) an assumption that players have conformity concerns over how the community perceives their private preferences, and (iii) an assumption that there is aggregate uncertainty about the distribution of private preferences in the population. I show that as all 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 preferences 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 preferences of their peers;

[30]Further, at this level of generality equilibria with partial revelation may exist and maximize the expected utility of the players.

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 preferences), they pool on an inefficient decision based on these imprecise beliefs. Finally, information about the fundamental state has a lower ability than information about peers' preferences to break a pooling equilibrium. My theoretical result that providing information about the preferences of one's peers is more effective than information about the fundamental state provides a framework to formalize intuitions extensively discussed empirically in social psychology and economics.

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chang, Edward H, Katherine L Milkman, Dolly Chugh, and Modupe Akinola, "Diversity thresholds: How social norms, visibility, and scrutiny relate to group composition," *Academy of Management Journal*, 2019, *62* (1), 144–171.

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

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

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

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

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

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

Fudenberg, Drew and Jean Tirole, "Perfect Bayesian equilibrium and sequential equilibrium," *journal of Economic Theory*, 1991, *53* (2), 236–260.

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

Giuliano, Paola and Nathan Nunn, "Understanding cultural persistence and change," *The Review of Economic Studies*, 2021, *88* (4), 1541–1581.

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.

Gulesci, Selim, Sam Jindani, Eliana La Ferrara, David Smerdon, Munshi Sulaiman, and H Young, "A stepping stone approach to norm transitions," 2023.

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

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

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

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

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

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

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

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

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

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

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

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

Rappoport, Daniel, "Reputational Delegation," 2022.

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.

Sheikh, Aziz, GBD 2016 Alcohol Collaborators et al., "Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016," *The Lancet*, 2018, *392* (10152), 1015.

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

_ , Peter Norman Sørensen, and Jianrong Tian, "Informational herding, optimal experimentation, and contrarianism," *The Review of Economic Studies*, 2021, *88* (5), 2527–2554.

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

_ , *How change happens*, Mit Press, 2019.

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

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

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

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

8 Appendix A

Proof of Proposition 1. Solving Equation (3) proves the existence of $\kappa^{\text{c.k.}}$ and the results when $\kappa \geq \kappa^{\text{c.k.}}$. Finally, the equilibrium with revelation is Pareto superior to the pooling equilibrium, which completes the proof. $\square$

Proof of Lemma 1. Fix a given period t and a conjectured linear belief $\alpha_t a_t + \beta_t$. The first-order condition given a conjectured belief of $\alpha_t a_t + \beta_t$ is equal to Equation (8). The normality assumption implies that the conditional expectation of both θ and μ will be linear in both v_t and s_t with an intercept. Using this observation, then for some exogenous constants, $c_{1,t}, \ldots, c_{6,t}$,

$$a_t(1+\kappa\alpha_t^2) = \kappa\alpha_t\beta_t + c_{1,t} + c_{2,t}\alpha_t\kappa + s_t(c_{3,t} + \kappa\alpha_t c_{4,t}) + v_t(1 + c_{5,t} + \kappa\alpha_t c_{6,t}). \tag{27}$$

Simplifying implies,

$$\frac{a_t(1+\kappa\alpha_t^2) - \kappa\alpha_t\beta_t - c_{1,t} - c_{2,t}\alpha_t\kappa}{c_{3,t} + \kappa\alpha_t c_{4,t}} = s_t + v_t\frac{1 + c_{5,t} + \kappa\alpha_t c_{6,t}}{c_{3,t} + \kappa\alpha_t c_{4,t}}. \tag{28}$$

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

$$\mathbf{E}\Big(v_t|h_t, a_t\Big) = c_{7,t} + c_{8,t}(\alpha_t, \kappa)\frac{a_t(1+\kappa\alpha_t^2) - \kappa\alpha_t\beta_t - c_{1,t} - c_{2,t}\alpha_t\kappa}{c_{3,t} + \kappa\alpha_t c_{4,t}}, \tag{29}$$

where $c_{8,t}(\alpha_t, \kappa) \neq 0$ is determined by both the prior beliefs and the conjectured equilibrium slope α_t. In equilibrium, the conjecture must be consistent implying the right-hand side of Equation (29) must equal $\alpha_t a_t + \beta_t$. This equality is stated below:

$$c_{8,t}(\alpha_t, \kappa)(1+\kappa\alpha_t^2) = \alpha_t(c_{3,t} + \kappa\alpha c_{4,t}) \tag{30}$$

$$c_{7,t}(c_{3,t} + \kappa\alpha_t c_{4,t}) + c_{8,t}(\alpha_t, \kappa)(-\kappa\alpha_t\beta_t - c_{1,t} - c_{2,t}\kappa\alpha_t) = \beta_t(c_{3,t} + \kappa\alpha_t c_{4,t}). \tag{31}$$

Note that for any solution to Equation (30), there exists a solution to the equation for β_t. This is because one can simplify Equation (31) to,

$$\beta_t\Big(c_{3,t} + \kappa\alpha_t c_{4,t} + \kappa\alpha_t c_{8,t}(\alpha_t, \kappa)\Big) = c_{7,t}(c_{3,t} + \kappa\alpha_t c_{4,t}) + c_{8,t}(\alpha_t, \kappa)(-c_{1,t} - c_{2,t}\kappa\alpha_t). \tag{32}$$

One can simplify the coefficient on β_t above as follows:

$$\begin{aligned} c_{3,t} + \kappa\alpha_t c_{4,t} + \kappa\alpha_t c_{8,t}(\alpha_t, \kappa) &= c_{8,t}(\alpha_t, \kappa)\frac{(1+\kappa\alpha_t^2)}{\alpha_t} + \kappa\alpha_t c_{8,t}(\alpha_t, \kappa) \\ &= c_{8,t}(\alpha_t, \kappa)(\frac{1}{\alpha_t} + 2\alpha_t\kappa), \end{aligned} \tag{33}$$

where the first equality comes from Equation (30). Finally, as $c_{8,t}(\alpha_t, \kappa) \neq 0$ and in any equilibrium with revelation $\alpha_t \neq 0$, a unique solution for β_t always exists.

Thus, the necessary and sufficient condition for an equilibrium with revelation is Equation (30). Finally, all of the terms in this equation are independent of the means of the prior beliefs ($c_{1,t}$ and $c_{2,t}$) and only condition on the precision matrix.

The proof of this first result implies that the realizations of y_t have no impact on whether or not beliefs converge, because in the Gaussian learning model, the realizations of y_t effect only the mean. Denote by τ_t the precision matrix of the beliefs at time t with the following parametrization.

$$\tau_t = \begin{pmatrix} \tau_{1,t} & \tau_{2,t} \\ \tau_{2,t} & \tau_{3,t} \end{pmatrix}. \tag{34}$$

The beliefs update as follows in any equilibrium with revelation,

$$\begin{aligned} \tau_{1,t+1} &= \tau_{1,t} + \phi_t \\ \tau_{3,t} &= \tau_{3,t} + 1 - \phi_t \\ \tau_{2,t} &= \tau_{2,t} + \sqrt{(1-\phi_t)\phi_t} \\ \phi_t &= \frac{(c_{3,t} + \kappa\alpha c_{4,t})^2}{(c_{3,t} + \kappa\alpha_t c_{4,t})^2 + (1 + c_{5,t} + \kappa\alpha_t c_{6,t})^2}. \end{aligned} \tag{35}$$

Further, $\theta(t) \to_p \theta$ if and only if

$$\lim \frac{\tau_{3,t}}{\tau_{1,t}\tau_{3,t} - \tau_{2,t}^2} \to 0, \tag{36}$$

namely the variance about θ converges to zero. If this variance converges to zero, however, then $c_{3,t}, c_{4,t}$, and $c_{5,t}$ (how s_t impacts the conditional expectation of θ, how s_t impacts the conditional expectation of μ, and how v_t impacts the conditional expectation of θ, respectively) all converge to zero. Further, $c_{6,t}$ (how v_t impacts the conditional expectation of μ) remains weakly positive. As a result, if $\theta(t) \to_p \theta$ and $\mu(t) \not\to_p \mu$, then $\phi_t \to 0$. However, if $\phi_t \to 0$ and there exist infinitely many periods with revelation, then $\tau_{3,t} >$

$\tau_{1,t} > 0$ implying

$$\lim \frac{\tau_{1,t}}{\tau_{1,t}\tau_{3,t} - \tau_{2,t}^2} \to 0, \tag{37}$$

namely the variance about μ converges to zero. This convergence implies $\mu(t) \to_p \mu$.

Similarly, if $\mu(t) \to_p \mu$, but $\theta(t) \not\to_p \theta$, then $c_{5,t}, c_{6,t}$ and $c_{4,t}$, but $c_{3,t}$ remains bounded away from zero. As a result, ϕ_t converges to a constant, $\phi \in (0,1)$. Hence,

$$0 < \lim \frac{\tau_{3,t}}{\tau_{1,t}\tau_{3,t} - \tau_{2,t}^2} < \lim \frac{(\phi + \epsilon)\tau_{1,t}}{\tau_{1,t}\tau_{3,t} - \tau_{2,t}^2}. \tag{38}$$

However, $\theta(t) \to_p \theta$ implies the outer limit converges to zero, and as a result so too does the inner limit. □

Proof of Lemma 2. I proceed by contradiction. Note that Lemma 1 implies $\theta(t) \to_p \theta \iff \mu(t) \to_p \mu$. Hence, we may suppose by contradiction that $\theta(t) \to_p \theta$ and $\mu(t) \to_p \mu$.

For the beliefs to converge, infinitely many periods of a decision rule with revelation must occur which is equivalent to Equation (30) holding for infinitely many periods. This implies (i) Equation (30) must hold as $t \to \infty$, $\theta(t) \to_p \theta$, and $\mu(t) \to_p \mu$. Further, (ii) because $\kappa > \kappa^{c.k.}$, there exists an $\epsilon > 0$ such that,

$$(1 + \kappa\alpha^2) - \alpha > \epsilon \quad \forall \alpha \geq 0. \tag{39}$$

Combining (i) and (ii), it must be that

$$\lim \frac{c_{8,t}(\alpha_t, \kappa)}{c_{3,t} + \kappa\alpha_t c_{4,t}} \not\to 1. \tag{40}$$

However, recall that this ratio in Equation (40) is defined by the following equation, as can be seen by manipulating Equation (28).

$$\begin{aligned} c_{9,t} + \frac{c_{8,t}(\alpha_t, \kappa)}{c_{3,t} + \kappa\alpha_t c_{4t}}_t (c_{3,t} + \kappa\alpha_t c_{4,t}) + v_t(1 + c_{5,t} + \kappa\alpha_t c_{6,t})\Big) \\ = \mathbf{E}(v_t | s_t(c_{3,t} + \kappa\alpha_t c_{4,t}) + v_t(1 + c_{5,t} + \kappa\alpha_t c_{6,t})), \end{aligned} \tag{41}$$

for some constant $c_{9,t}$. Finally, because the beliefs converge, then $c_{3,t}, c_{4,t}, c_{5,t}$, and $c_{6,t}$ converge to zero implying that the right-hand side converges to $\mathbf{E}(v_t|v_t)$. This implies that the limit in Equation (40) does converge to 1, which derives our contradiction. □

Proof of Lemma 3. It is sufficient to show that in every period the equilibrium involves revelation. To do so, we must show that there exists a solution to Equation (30). First, one can show that as $\alpha_t \to \infty$ the left-hand side is greater than the right-hand side, implying that a sufficient condition is that there exists conjectured beliefs where,

$$c_{8,t}(\alpha_t, \kappa)(1 + \kappa\alpha_t^2) < \alpha_t(c_{3,t} + \kappa\alpha_t c_{4,t}). \tag{42}$$

Further, as $\kappa < \kappa^{\text{c.k.}}$, upon setting $\alpha_t = 2$ implies that a sufficient condition is

$$\frac{c_{8,t}(2, \kappa)}{c_{3,t} + 2\kappa c_{4,t}} \leq 1 + \epsilon. \tag{43}$$

However, for any ϵ this inequality holds for sufficiently precise beliefs about θ and μ, because (i) the left-hand side is continuous with respect to the variance matrix of the public beliefs about θ, v, (ii) the left-hand side is equal to one when the variance matrix is equal to zero. (i) and (ii) imply that there exists sufficiently precise beliefs where such a condition holds for not only those beliefs but any beliefs more precise. □

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

The proof of statement (1) is as follows. I claim that $\tilde{s}(0, 0, \theta(t), \mu(t)) < 1 - \epsilon$ is a sufficient condition, where $\tilde{s}(\cdot)$ is as defined in Equation (10). This condition is sufficient as (i) the left-hand side of Equation (10) is larger as $\alpha \to \infty$ (ii) that such a solution holds when $\kappa = 0$, and (iii) the condition is continuous with respect to κ.

When $\kappa = \alpha = 0$, Equation (8), simplifies to

$$\gamma_t \mathbf{E}(\theta | \theta(t), \mu(t), s_t, v_t) + v_t \tag{44}$$

Further, given the definition of $\tilde{s}(\alpha, \kappa, \theta(t), \mu(t))$ in Equation (9), Equation (44) has a sensitivity strictly less than $1 - \epsilon$ for any beliefs concluding the proof.

The proof of the final statement, whereby there exist an open set of parameter values where the beliefs do not converge, despite $\kappa < \kappa^{\text{c.k.}}$ can be seen by noting that Equation (30) is cubic in α_t. One can further show that because the coefficient on the term α_t^3 is negative the signalling equilibrium involves revelation if and only if there exists a positive root. Further, by Descartes rule of signs, there may be either zero or two positive roots. Hence, the necessary and sufficient condition for a positive solution to exist is for there to exist three roots. Further, for any cubic polynomial $A + Bx + Cx^2 + Dx^3$, there exist three real roots if and only if

$$(-27 \cdot D^2 \cdot A^2 + 18 \cdot B \cdot C \cdot D \cdot A - 4 \cdot D \cdot B^3 - 4 \cdot C^3 \cdot A + B^2 \cdot C^2) \geq 0. \tag{45}$$

Hence it suffices to show that the Equation (45) is strictly negative despite $\kappa < \kappa^{\text{c.k.}}$. As such an expression is continuous, demonstrating that this expression may be strictly negative will prove the results. Further, upon simplifying this expression one can find a precision matrix of the form in Equation (34), where Equation (45) is negative when $\gamma = 1$ and $\kappa = .9\kappa^{\text{c.k.}}$.[31] □

Proof of Proposition 3. This analysis uses τ_ϵ and $\tau_{v_{\text{ind}}}$ to denote the precision of a player's signal and precision of a player's type, which were normalized to one in the main analysis by assuming that the variance of ϵ_t and ν_t were equal to one. A linear decision rule can be written as,

$$
\begin{aligned}
a_t &= \beta_t \bar{\theta}(t) + \alpha_t \big(v_t + \mathbf{E}(\theta | \theta(t), s_t) - \bar{\theta}(t)\big). \\
\frac{a_t - \beta_t \bar{\theta}(t)}{\alpha_t} &= \big(\mathbf{E}(\theta) - \bar{\theta}(t)\big) + v_t \\
&= \big(\frac{\tau_{\theta,t}\bar{\theta}(t) + \tau_\epsilon s_t}{\tau_{\theta,t} + \tau_\epsilon} - \bar{\theta}(t)\big) + v_t \\
&= \frac{\tau_\epsilon(\theta - \bar{\theta}(t) + \epsilon_t)}{\tau_{\theta,t} + \tau_\epsilon} + v_t, \qquad (46)
\end{aligned}
$$

for some slope α_t and intercept β_t. Note that the left-hand side can be computed given a_t and yields v_t plus a Gaussian random variable $\frac{\tau_\epsilon(\theta - \bar{\theta}(t) + \epsilon_t)}{\tau_{\theta,t} + \tau_\epsilon}$. Next, note that $\theta - \bar{\theta}(t)$ and ϵ_t are mean zero, hence the random variable $(a_t - \beta_t \bar{\theta}(t))/\alpha_t - v_t$ is mean zero. Further, its precision is,

$$
\big(\frac{\tau_{\theta,t} + \tau_\epsilon}{\tau_\epsilon}\big)^2 \big(\frac{\tau_{\theta,t}\tau_\epsilon}{\tau_{\theta,t} + \tau_\epsilon}\big) = \frac{\tau_{\theta,t}(\tau_{\theta,t} + \tau_\epsilon)}{\tau_\epsilon}, \qquad (47)
$$

as the first term notes $\theta - \bar{\theta}(t) + \epsilon_t$ is scaled by $\frac{\tau_\epsilon}{\tau_\epsilon + \tau_{\theta,t}}$. Hence, the precision is scaled by the reciprocal of such a constant squared and the second term is the precision of $\theta - \bar{\theta}(t) + \epsilon_t$. Therefore,

$$
\hat{v}_t = \frac{a_t - \beta_t \bar{\theta}(t)}{\alpha_t} \frac{\tau_{\theta,t}(\tau_{\theta,t} + \tau_\epsilon)}{\tau_{\theta,t}(\tau_{\theta,t} + \tau_\epsilon) + \tau_{v_{\text{ind}}}\tau_\epsilon} := \frac{a_t - \beta_t \bar{\theta}(t)}{\alpha_t} \tilde{\tau}_t. \qquad (48)
$$

[31] Mathematica code available upon request.

Given $\hat{v}_t$, one can compute the first-order condition for player t:

$$\begin{aligned}
a_t - \mathbf{E}(\theta) - v_t + \kappa\frac{(a_t - \beta_t\bar{\theta}(t))\tilde{\tau}_t^2}{\alpha_t^2} &= 0 \\
\iff a_t(1 + \frac{\kappa\tilde{\tau}_t^2}{\alpha_t^2}) &= \mathbf{E}(\theta) + v_t + \kappa\frac{\beta_t\bar{\theta}(t)\tilde{\tau}_t^2}{\alpha_t^2} \\
&= \bar{\theta}(t) + \frac{\mathbf{E}(s_t - \bar{\theta}(t))}{\tau_{\theta,t} + \tau_\epsilon} + v_t + \kappa\frac{\beta_t\bar{\theta}(t)\tilde{\tau}_t^2}{\alpha_t^2} \\
&= \bar{\theta}(t)(1 + \kappa\frac{\beta_t\tilde{\tau}_t^2}{\alpha_t^2}) + v_t + \mathbf{E}(\theta) - \bar{\theta}(t). \qquad (49)
\end{aligned}$$

Thus,

$$a_t = \bar{\theta}(t)\frac{1 + \kappa\frac{\beta_t\tilde{\tau}_t^2}{\alpha_t^2}}{1 + \frac{\kappa\tilde{\tau}_t^2}{\alpha_t^2}} + \frac{v_t + \mathbf{E}(\theta) - \bar{\theta}(t)}{1 + \frac{\kappa\tilde{\tau}_t^2}{\alpha_t^2}}. \qquad (50)$$

Finally, the conjecture must be correct in equilibrium implying,

$$\beta_t = \frac{1 + \kappa\frac{\beta_t\tilde{\tau}_t^2}{\alpha_t^2}}{1 + \frac{\kappa\tilde{\tau}_t^2}{\alpha_t^2}} \iff \beta_t = 1 \qquad (51)$$

$$\alpha_t = \frac{1}{1 + \frac{\kappa\tilde{\tau}_t^2}{\alpha_t^2}} \iff \alpha_t^2 - \alpha_t + \kappa\tilde{\tau}_t^2 = 0. \qquad (52)$$

Given this decision rule, an equilibrium with revelation exists if and only if $\kappa\tilde{\tau}_t \leq 1/4$. Further, $\tilde{\tau}_t$ is monotone in $\tau_{\theta,t}$ which implies that social learning about fundamentals occurs if and only if $\kappa \leq \kappa^{\text{c.k.}}$. When $\kappa > \kappa^{\text{c.k.}}$ the players switch to a pooling equilibrium when $\kappa\tilde{\tau}_t < 1/4$. As a result, fixing any prior beliefs, the limit of $\tilde{\tau}_t$ is monotone decreasing in κ, implying the limit of $\tau_{\theta,t}(\kappa)$ is monotone decreasing in κ.

I now prove the comparative statics regarding the asymptotic adaptation loss and utility. When $\kappa \leq \kappa^{\text{c.k.}}$, the asymptotic adaptation loss and utility converges to that of the common knowledge benchmark where the comparative statics with respect to κ were already established. When $\kappa > \kappa^{\text{c.k.}}$ the adaptation loss, in the limit, is equal to,

$$\mathbf{E}(\theta + v_t - \bar{\theta}(t))^2 = \mathbf{E}(v_t^2) + \frac{1}{\tau_{\theta,t}}. \qquad (53)$$

Further, given the comparative static with respect to κ of $\tau_\theta(\kappa)$, it follows that the adaptation loss is decreasing in κ. Finally, the asymptotic utility is equal to the asymptotic adaptation plus the conformity loss, and the conformity loss is mechanically decreasing in

κ which proves the result. $\square$

Proof of Proposition 4. The proof of the first statement is a direct consequence of the proof of statement 2 of Proposition 2.

The proof of the second statement can be seen by the third statement of Proposition 2 when n is finite. If n is infinite, then the proof follows from either of the uni-demnsional uncertainty analyses in Proposition 3 when $\rho = 0$ or Proposition 7 when $\delta = 0$.

The proof of the third statement contains two parts. The proof that a pooling equilibrium may be fragile to a peer-oriented intervention is provided in Figure 1 which shows that an increase in $\tau_{\mu,t}$ may break a pool. The proof that an increase in $\tau_{\theta,t}$ never breaks a pooling equilibrium can be seen by noting that when there is no correlation in the beliefs of θ and μ, Equation (8) reduces to

$$a_t(1 + \kappa\alpha^2) = \gamma_t \mathbf{E}\big(\theta \mid \theta(t), s_t\big) + v_t + \kappa\alpha\beta + \kappa\alpha\mathbf{E}\big(\mu \mid \mu(t), v_t\big) \tag{54}$$

due to independence. One can now see that uncertainty about θ decreases $\tilde{s}(\alpha, \kappa, \theta(t), \mu(t))$ in Equation (10). Finally, as argued in Lemma 3 as $\alpha \to \infty$ the left-hand side is greater than the right-hand side, hence the necessary and sufficient condition is whether there exists an α such that the left-hand side is less than the right-hand side. As $\tilde{s}(\alpha, \kappa, \theta(t), \mu(t))$ uniformly decreases when uncertainty about θ is decreased, one can see that a decrease in uncertainty about θ will never break a pooling equilibrium. $\square$

Proof of Proposition 5. The players pool on $a^* = \mathbf{E}(\theta) + \mathbf{E}(\mu)$ where both expectations are derived from the decision of player 1. Player 1 chooses a_1 as follows:

$$\begin{aligned} a_1 &= \lambda_\theta \mathbf{E}(\theta | s_1) + \lambda_v \mathbf{E}(\mu | v_1) && (55) \\ &= \lambda_\theta\big(\theta + N(0, 1 + \tau_\theta)\big) + \lambda_v\big(\mu + N(0, 1 + \tau_\mu)\big) && (56) \end{aligned}$$

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

$$\frac{a_1 - \lambda_\theta\theta}{\lambda_v} = \mu + N(0, 1 + \tau_\mu) + \frac{\lambda_\theta}{\lambda_v} N(0, 1 + \tau_\theta). \tag{57}$$

As a result,

$$\mathbf{E}(\mu \mid a_1) = \frac{\frac{a_1 - \lambda_\theta\theta}{\lambda_v}\left(\frac{1}{1+\tau_\mu} + (\frac{\lambda_\theta}{\lambda_v})^2 \frac{1}{1+\tau_\theta}\right)^{-1}}{\tau_\mu + \left(\frac{1}{1+\tau_\mu} + (\frac{\lambda_\theta}{\lambda_v})^2 \frac{1}{1+\tau_\theta}\right)^{-1}} = \frac{a_1 - \lambda_\theta\theta}{\lambda_v} \frac{1}{1 + \tau_\mu\left(\frac{1}{1+\tau_\mu} + (\frac{\lambda_\theta}{\lambda_v})^2 \frac{1}{1+\tau_\theta}\right)}. \tag{58}$$

Simplifying the posterior expectation of μ produces the result in the proposition. Further, a symmetric calculation occurs when considering an individual-oriented intervention. $\square$

Proof of Proposition 6. Note that when θ is common knowledge the condition for the existence of a revealing equilibrium reduces to a function $\kappa^*(\tau_{\mu,t})$ whereby a decision rule with revelation exists if and only if $\kappa < \kappa^*(\tau_{\mu,t})$ (for a proof, set $\delta = 0$ in the characterization in Proposition 7).

The threshold κ^* is determined by the value of κ that binds Equation (18) when $\tau_{\mu,t} = \max_{t'} \tau_{\mu,t'} := \tau^*$ in an equilibrium where every period involves revelation. As $\tau_{\mu,t} < \tau_\mu/\rho^2$ and $\rho > 0$, then $\kappa^* > \kappa^{\text{c.k.}}$.

Statement (1) and (3) follow immediately from the criterion for when the decision rule in period t involves revelation.

Finally, statement (2) follows from noting that when $\kappa > \kappa^*$, the equilibrium must have infinitely many periods of pooling. If the equilibrium had finitely many periods of pooling, then consider the final period of pooling. After this period, there will exist a period where the precision is τ^*. Further, as $\kappa > \kappa^*$, the decision rule must involve revelation in this period deriving a contradiction. Similarly, if there are only finitely many periods of revelation, then after all the periods of pooling, $\tau_{\mu,t} \to \tau_\mu$. However, at this point, the decision rule must involve revelation. Finally, that there are infinitely periods of pooling and infinitely many periods of revelation implies $\tau_{\mu,t}$ is not Cauchy and thus does not converge. $\square$

Proof of Proposition 7. As in the proof of Proposition 3, I will denote by $\tau_{v_{\text{ind}}}$ the precision of ν_t which was normalized to one, where ν_t was defined to satisfy $v_t := \mu + \nu_t$. Further, sufficient statistics for the decision rule in period t are the current mean and precision of $\mu(t)$ which are denoted as $\bar{\mu}(t)$ and $\tau_{\mu,t}$. I will now consider whether the decision rule for player i can involve revelation in period $t(i)$, which will be denoted as t for the rest of the proof. A linear equilibrium with revelation, when it exists, is of the form,

$$a_t = \alpha_t v_i + \beta_t. \tag{59}$$

The public perception of v_i given a_t is equal to $(a_t - \beta_t)/\alpha_t$. Therefore, one can write the utility as a function of a_t as follows,

$$\begin{aligned} u_i(a_t) = &-(a_t - v_i)^2 - \kappa\Big(\frac{a_t - \beta_t}{\alpha_t} - \frac{\bar{\mu}(t)\tau_{\mu,t} + v_i\tau_{v_{\text{ind}}}}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}}\Big)^2 \\ &- \frac{\delta}{1-\delta}\big(\frac{a_t - \beta_t}{\alpha_t} - v_i\big)^2 - \frac{\delta}{1-\delta}\kappa\Big(\frac{a_t - \beta_t}{\alpha_t} - \frac{\bar{\mu}(t)\tau_{\mu,t} + v_i\tau_{v_{\text{ind}}}}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}}\Big)^2, \end{aligned} \tag{60}$$

where the first line denotes the payoffs in period t given player i's perception of μ and the second line denotes the continuation payoff in all subsequent periods. One can now take a first-order condition to generate:

$$0 = a_t - v_i + \frac{\kappa}{\alpha_t}\frac{1}{1-\delta}\Big(\frac{a_t - \beta_t}{\alpha_t} - \frac{\bar{\mu}(t)\tau_{\mu,t} + v_i\tau_{v_{\text{ind}}}}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}}\Big) + \alpha_t\frac{\delta}{1-\delta}\big(\frac{a_t - \beta_t}{\alpha_t} - v_i\big). \tag{61}$$

Finally, in equilibrium, the beliefs are correct which implies $v_t = (a_t - \beta_t)/\alpha_t$. Simplifying the remaining terms implies:

$$\begin{aligned} 0 &= a_t - v_i + \frac{\kappa}{\alpha_t}\frac{1}{1-\delta}\Big(v_i - \frac{\bar{\mu}(t)\tau_{\mu,t} + v_i\tau_{v_{\text{ind}}}}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}}\Big) \\ &= a_t - v_i + \frac{\kappa}{\alpha_t}\frac{1}{1-\delta}\frac{v_i\tau_{\mu,t} - \bar{\mu}(t)\tau_{\mu,t}}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}} \\ \iff a_t &= v_i\big(1 - \frac{\kappa}{\alpha_t}\frac{1}{1-\delta}\frac{\tau_{\mu,t}}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}}\big) + \frac{\kappa}{\alpha_t}\frac{1}{1-\delta}\frac{\tau_{\mu,t}\bar{\mu}(t)}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}}. \end{aligned} \tag{62}$$

In equilibrium, α_t must equal the coefficient on v_i in the decision rule:

$$\alpha_t = 1 - \frac{\kappa}{\alpha_t}\frac{1}{1-\delta}\frac{\tau_{\mu,t}}{\tau_{\mu,t} + \tau_{v_{\text{ind}}}} \iff \alpha_t = \frac{1 \pm \sqrt{1 - 4\kappa\frac{1}{1-\delta}\frac{\tau_{\mu,t}}{\tau_{\mu,t}+\tau_{v_{\text{ind}}}}}}{2}. \tag{63}$$

Further, note that given α_t, the solution to β_t is uniquely determined by Equation (62). Hence, a solution exists if and only if the condition in the text holds because this condition corresponds to the term in the square root being non-negative. □

Proof of Lemma 4. This is precisely Theorem 3 in Bernheim (1994). □

Proof of Proposition 8. The proof of this Proposition is a direct consequence of Proposition 11 in Appendix B. □

Proof of Proposition 9. First note that by an identical argument to Bernheim (1994) all equilibria must be central pooling. Given a central pooling equilibrium, Equation (24) characterizes the equilibrium outside of the central pool.

Let us begin showing that if κ exceeds the threshold in the proposition no solution to the differential equation exists for any initial condition, and, as a direct consequence, no equilibrium with revelation exists. To do so, one can differentiate Equation (24) and determine,

$$a_t''(v_t) = \frac{-\kappa\big(v_t - g_t(v_t)\big)\big(1 - a_t'(v_t)\big)}{\big(v_t - a_t(v_t)\big)^2} + \frac{\kappa\big(1 - g_t(v_t)\big)}{v_t - a_t(v_t)}$$

$$\iff \big(v_t - g_t(v_t)\big)a_t''(v_t) = a_t'(v_t)\left(1 - g_t'(v_t) - \frac{a_t'(v_t)\big(1 - a_t'(v_t)\big)}{\kappa}\right). \tag{64}$$

Further, for any value x, $x(1-x)/\kappa < (1/4)/\kappa$ and given the condition on $\sup_x g_t'(x)$, one can show that for any κ greater than $1/4(1-s_t)$ there exists an $\epsilon_3 > 0$ such that,

$$\big(v_t - g_t(v_t)\big)a_t''(v_t) \geq a_t'(v_t)\epsilon_3 \iff a_t''(v_t) \geq \frac{a_t'(v_t)\epsilon_3}{v_t}, \tag{65}$$

where the final inequality comes from noting that $g_t(v_t) \geq 0$ for positive values of v_t. Let $a_0(v_t)$ be the decision rule that binds differential inequality in Equation (65) and satisfies $a_t(\bar{v}_t) = a_0(\bar{v}_t)$, where $\bar{v}_t$ denotes the supremum of the central pool.

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

$$a_0(v_t) = c_1 v_t^{1+\epsilon_3} + c_2, \tag{66}$$

where $c_1 > 0$. However, one can show that

$$a_t(v_t) = a_t(\bar{v}_t) + \int\int a_t''(v_t) \geq a_0(\bar{v}_t) + \int\int a_0''(v_t) = c_1 v_t^{1+\epsilon_3} + c_2. \tag{67}$$

This inequality implies that $a_t(v_t) > v_t$ for a positive value of v_t which is a contradiction because the players always have an incentive to choose a mildly more conforming decision and receive both a better adaptation loss and a better conformity loss.

Now I will show that if κ is less than the condition provided in Proposition 9 an equilibrium with full revelation exists. To do so, first note that there always exists a v_t such that $v_t = g_t(v_t)$, by Bayes Plausability. To see why, note that if there existed two values v_t' and v_t'' such that $v_t' > g_t(v_t')$ and $v_t'' < g_t(v_t'')$ then by continuity there exists a value v_t where $v_t = g_t(v_t)$. So for no solution to exist, up to symmetry, it must be the case that $v_t' > g_t(v_t')$ for all values of v_t. By monotonicity of the integral, $\mathbf{E}(v_t) > \mathbf{E}\big(g_t(v_t')\big) = \mathbf{E}(\mu)$ which is a contradiction of Bayes Plausability. As a similar argument could be made if instead $v_t' < g_t(v_t')$ for all values, there must exist a value v_t^* such that $v_t^* = g_t(v_t^*)$. I will now show that if κ satisfies the condition in the proposition there exists an equilibrium with full revelation where $a_t(v_t^*) = v_t^*$. To do so, I will use Carathéodory's existence theorem.

To be able to apply this theorem to Equation (24) the decision rule, $a_t(v_t)$ must never cross v_t except for v_t^* so that the implicit function in Equation (24) is continuous on its domain. I will analyze the differential equation to the right of v_t^* and a symmetric analysis occurs to the left. A sufficient condition for $a_t(v_t) < v_t$ is that $a_t'(v_t) \leq 1/2$ for all $v_t \geq v_t^*$. Taking Equation (24), this condition can be stated as

$$\begin{aligned} & 2\kappa\big(v_t - g_t(v_t)\big) \leq \frac{1}{2}\big(v_t - a_t(v_t)\big) \\ \Longleftarrow\ & 2\kappa\big(v_t - g_t(v_t)\big) \leq \frac{1}{2}\big(v_t - v_t^*\big) \\ \Longleftrightarrow\ & \kappa \leq \frac{1}{4}\frac{1}{1 - \frac{g_t(v_t) - v_t^*}{v_t - v_t^*}} \end{aligned} \tag{68}$$

Finally, the expression in the denominator of Equation (68) can be simplified as follows to generate the condition in the proposition.

$$\frac{g_t(v_t) - v_t^*}{v_t - v_t^*} \geq \frac{g_t(v_t^*) + i_t(v_t - v_t^*) - v_t^*}{v_t - v_t^*} = i_t \tag{69}$$

Applying this inequality to Equation (68) finishes the proof. □

9 Appendix B: Online

This section details extensions then analyzes the equilibria that satisfy D1 and Pareto-Optimality. I begin this analysis when θ, μ are common knowledge. Such an analysis holds for a general class of distributions. Next, I detail how one can extend the results to allow for learning about these two random variables.

9.1 Extensions

Let us begin with an analysis with conformist preferences where individuals want their perceived type to be b units higher than the average in the population. Further, I consider the environment of Section 3 without uncertainty allowing us to drop any time-dependence. A player with type v's utility is,

$$-(a - v)^2 - \kappa\big(\hat{v}(a) + b\big)^2. \tag{70}$$

Proposition 10 (**Shifted Preferences**)
Suppose the preferences of the players are as in Equation (70)*, then in the linear equilibrium where $\hat{v}(a) = \alpha \cdot a + \beta$, α is independent of b.*

This result shows that while choosing a value $b \neq 0$ will lead to different equilibrium decisions, the level of adaptiveness remains the same. This result can easily be generalized to show that the results regarding social learning are not sensitive to this exact parametrization of conformity concerns.

Proof of Proposition 10. One can conjecture equilibrium beliefs of $\hat{v}(a) = \alpha a + \beta$. Given these conjectures the first-order condition of a player is,

$$a - v + \kappa\alpha(\alpha a + \beta + b) = 0 \iff v = (1 + \kappa\alpha^2)a + \kappa\alpha(\beta + b). \tag{71}$$

In equilibrium the conjectures are correct which implies,

$$\alpha = 1 + \kappa\alpha^2 \text{ and } \beta = \kappa\alpha(\beta + b). \tag{72}$$

One can note the solution to α is independent of b, and if a solution for α exists, there will always exist a solution for β. □

9.2 Sufficiency of Linear Equilibria

Throughout this subsection the only distributional assumptions that are needed about v are that (i) the distribution admits a continuous density and (ii) the support of the distribution is the real line.

Let us first begin with equilibria that satisfy D1. As detailed in the text, Bernheim (1994) implies that Equation (8) must hold outside the central pool. Given this result, one can show the following Proposition.

Proposition 11 (**Linear Equilibria**)
Any equilibrium that satisfies D1 has an empty central pool. On either side of the empty central pool, the decision rule is characterized by $a = \alpha v$ where α takes one of two possible values: $\frac{1-\sqrt{1-4\kappa}}{2}, \frac{1+\sqrt{1-4\kappa}}{2}$ which exist if and only if $\kappa \leq 1/4$, where κ denotes the weight on conformity. However, there always exists a fully-pooling decision rule in which all types take the same decision.

Proof of Proposition 11. As mentioned in the text, there always exists the fully pooling decision rule where the central pool is the entire domain. Hence, let us consider central pooling equilibria with central pools that are not the entire domain. Without loss of generality let us assume $\bar{v} < \infty$, and an identical characterization follows if $\underline{v} > -\infty$. The first-order condition implies:

$$a - v + \kappa \hat{v}(a)\hat{v}'(a) = 0. \tag{73}$$

Substituting $a = x$ and $ay(x) = \hat{v}(x)$ to Equation (73) yields,

$$0 = x - xy(x) + \kappa xy(x)\big(y(x) + x\dot{y}(x)\big) \iff \dot{y}(x) = \frac{-1 + y(x) - \kappa y(x)^2}{\kappa xy(x)}. \tag{74}$$

This simplification is well defined because the denominator is non-zero for all x outside the central pool. Next, if the equilibrium decision rule is linear, then $\dot{y}(x) = 0 \ \forall \ x$. Further, if $\dot{y}(x) = 0$ for some x, one can calculate the $\ddot{y}(\cdot)$ as,

$$\begin{aligned} \ddot{y}(x) = \frac{-y(x)^2(kx\dot{y}(x) + 1) + ky(x)^3 + x\dot{y}(x) + y(x)}{kx^2y(x)^2} &= \frac{-y(x)^2 + ky(x)^3 + y(x)}{kx^2y(x)^2} \\ &= -\frac{\dot{y}(x)}{x} = 0. \end{aligned} \tag{75}$$

Equation (75) implies if $\dot{y}(x) = 0$ for any x, then $\dot{y}(x) = 0 \ \forall \ x$. Thus any non-linear solution satisfies the following integral expression derived by re-arranging Equation (74):

$$\begin{aligned} &\int \frac{\dot{y}(x)\kappa y(x)}{-1 + y(x) - \kappa y(x)^2} = \int \frac{-1}{x} \iff \\ &\frac{-\log(1 - y(x) + \kappa y(x)^2)}{2} + \frac{\log(1 + \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}) - \log(1 - \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}})}{2\sqrt{1-4\kappa}} + c = -\log(x), \end{aligned} \tag{76}$$

where c is the constant of integration. If $\kappa = 1/4$, then Equation (76) is ill-defined and thus all solutions are necessarily linear. Further, assuming $\kappa < 1/4$, any non-linear solution satisfies the following equalities:

$$\log\left(\sqrt{1 - y(x) + \kappa y(x)^2} \cdot \Big(\frac{1 - \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}{1 + \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}\Big)^{\frac{1}{2\sqrt{1-4\kappa}}}\right) = \log(e^c x) \tag{77}$$

$$\iff \sqrt{1 - y(x) + \kappa y(x)^2} \cdot \Big(\frac{1 - \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}{1 + \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}\Big)^{\frac{1}{2\sqrt{1-4\kappa}}} = e^c x \tag{78}$$

$$\iff (1 - y(x) + \kappa y(x)^2)\Big(\frac{1 - \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}{1 + \frac{1+2\kappa y(x)}{\sqrt{1-4\kappa}}}\Big)^{\frac{1}{\sqrt{1-4\kappa}}} = x^2 e^c. \tag{79}$$

Recall that the map $a(v)$ is defined on $(\bar{v}, \infty)$ and thus one can consider the limit of the above equation as $x \to \infty$. The right-hand side of the equation diverges as $x \to \infty$ implying the left-hand side must also diverge. However, if $y(x)$ is bounded, then both the first and second terms of the left-hand side will remain bounded. Therefore, one can assume that $y(x)$ is unbounded. By definition, then $v/a(v)$ is unbounded. However, if $v/a(v)$ is unbounded, then there must exist a player for which ϵ less conformity is preferred.

Thus no non-linear solution exists when $\kappa < 1/4$. If instead $\kappa > 1/4$, then the simplification to Equation (76) is

$$\frac{1}{2}\log\left(1 - y(x) + \kappa y(x)^2\right) + \frac{\tan^{-1}\left(\frac{2ky(x)-1}{\sqrt{4k-1}}\right)}{\sqrt{4k-1}} + c = \log(x). \tag{80}$$

As argued above, $y(x)$ must be bounded. However, this implies the left-hand side of the above equation cannot diverge, but the right-hand side necessarily diverges. As no non-linear solution exists for any κ, the only solutions to the differential equation outside the central pool are linear. I will now show that the central pool is empty.

Equation (3) shows that the linear equilibria must satisfy:

$$\alpha(1-\alpha) = \kappa \iff \alpha = \frac{1 \pm \sqrt{1-4\kappa}}{2}. \tag{81}$$

Thus for a central pooling equilibrium to exist one must find $a^*, \underline{v}, \bar{v}$ such that the player with preferences $\bar{v}$ is indifferent between a^* and $\alpha\bar{v}$. I consider the case where $a^* \leq 0$ and an identical argument holds if $a^* \geq 0$.[32] Note that the conformity loss following a^* is at best zero. Fixing the slope of the linear decision rule $d(v) = \alpha v$, then the following inequality must hold:

$$-(\bar{v} - a^*)^2 \geq -(1-\alpha)^2\bar{v}^2 - \kappa\bar{v}^2, \tag{82}$$

where the left-hand side is an upper bound on the utility in the central pool and the right-hand side is the utility in the linear equilibrium. One can notice that $\bar{v}^2$ cancels out and that for the solutions to α such an inequality never holds. Given that such an inequality cannot hold, the central pool must be empty in any equilibrium with revelation. □

While Proposition 11 implies that multiple equilibria with revelation exist, only one equilibrium is Pareto optimal. Note that the conformity loss is identical across the equilibria by Bayes Plausibility. Hence, the equilibrium that maximizes utility is the one that minimizes the adaptation loss. In the class of linear equilibria, this corresponds to the

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

equilibrium with the highest slope (as all the slopes are strictly less than one). These refinements select a unique equilibrium as stated in Proposition 1 in the text.

9.3 Generalized Distributions

Subsection 6.3 analyzed the impact of conformity concerns on social learning about preferences when θ was common knowledge for a general class of distributions. This subsection will analyze the impact of conformity concerns on social learning about fundamentals when μ is common knowledge for a general class of distributions.

Suppose each player has preferences $v_t \sim F$ where F admits a continuous and differentiable density $f(\cdot)$ with support equal to the real line and mean normalized to zero. Further, denote by ω_t the player t's posterior beliefs about θ in period t after observing s_t. I will assume the distribution of ω_t admits a continuous and differentiable density. The only decision relevant information for the player is $v_t + \omega_t := \tilde{v}_t$ and it is without loss to consider equilibria that only condition on the sum. Further, by an identical argument to Bernheim (1994), the only equilibria satisfying D1 are central pooling equilibria as a function of $\tilde{v}_t$.

Let us detail the conformity concerns in this general framework. As previously noted, outside the central pool the decision rule will be strictly increasing in $\tilde{v}_t$ which implies the existence of an inverse. Denote by x an arbitrary decision where $a^{-1}(x) = \tilde{v}_t$, then the conformity loss upon choosing the decision x is

$$\kappa\Big(\tilde{v}_t - h_1\big(\tilde{v}_t\big)\Big)^2 + \kappa h_2\big(\tilde{v}_t\big). \tag{83}$$

Note that a player has concerns over the second moment of their perceived preferences. This second moment has two components the mean and the variance. The mean of the perceived preferences is simply $\tilde{v}_t$ less $h_1\big(\tilde{v}_t\big)$, where $h_1(t)$ is defined such to make such an equality hold. Without uncertainty, $h_1(\tilde{v}_t) = 0$, however, in general $h_1(\tilde{v}_t) \neq 0$ because a high value of $\tilde{v}_t$ may stem from either a high value of v_t or a high value of ω_t. Further, $h_2(\tilde{v}_t)$ represents the impact of the variance of players with different preferences choosing the same decision on the conformity loss. Formally, $h_1\big(\tilde{v}_t\big) = \tilde{v}_t - \mathbf{E}(v_t | v_t + \omega_t = \tilde{v}_t)$ and $h_2\big(v_t | v_t + \omega_t = \tilde{v}_t)\big) = \mathbf{Var}(v_t | v_t + \omega_t = \tilde{v}_t)$.

Rather than dealing with the distributions of v_t and ω_t, I will work with $h_1(\cdot)$ and $h_2(\cdot)$.

Differentiating the utility function in Equation (83) implies:

$$0 = 2(a_t - \tilde{v}_t) + 2\kappa\Big(\tilde{v}_t - h_1(\tilde{v}_t)\Big)\frac{1}{a'(\tilde{v}_t)}\big(1 - h_1'(\tilde{v}_t)\big) + \frac{\kappa h_2'(\tilde{v}_t)}{a'(\tilde{v}_t)}$$

$$\iff a'(v_t) = \frac{\kappa\Big(\big(\tilde{v}_t - h_1(\tilde{v}_t)\big)\big(1 - h_1'(\tilde{v}_t)\big) + \frac{1}{2}h_2'(\tilde{v}_t)\Big)}{a_t - v_t}. \tag{84}$$

However, such an expression has a similar formulation to Equation (24) in Proposition 9. One can replace $g_t(v_t)$ in Equation (24) with the expression in Equation (84).

9.4 Generalized Results With Social Learning

The first two subsubsections show that the linearity is implied by D1 when either μ or θ are common knowledge, respectively. The third subsubsection discusses the difficulty with both dimensions of uncertainty. The final subsubsection discusses the learning outcomes if $\gamma_t = 1$ in all periods.

9.4.1 Learning About the Fundamental State:

In this extension, I assume μ is common knowledge, and without loss of generality equal to zero. By induction, it suffices to consider the incentives of an arbitrary player and thus drop any time dependence. Here, $\theta \sim N(0, \tau_\theta)$ and $v \sim N(0, \tau_\mu)$, resulting in the following utility for player 1:

$$-(a - \theta - v)^2 - \kappa \int \phi(b, a) b^2 db. \tag{85}$$

Proposition 12 (**Linear Equilibria without Population Uncertainty**)
The only equilibria with revelation that satisfy D1 are the linear equilibria.

Proof of Proposition 12. As $\theta + v$ is the only decision-relevant variable for the players it is without loss to consider decision rules such that a is a function of the sum alone. Note that in any equilibrium in which a is a function of $\theta + v$, an identical proof to Bernheim (1994) can be used to show that D1 implies central pooling. Finally, note that outside the central pool if $a(x) = a^*$, then the equilibrium inference given a^* is a Gaussian random variable with mean $\frac{\tau_\theta}{\tau_\theta + \tau_\mu} x$ and a fixed variance. Given the quadratic loss framework the variance is irrelevant in determining which decision to take outside the central pool. Hence, the decision outside the central pool is chosen to maximize,

$$-(a - v - \theta)^2 - \kappa \frac{\tau_\theta}{\tau_\theta + \tau_\mu} \tilde{a}^{-1}(a)^2, \tag{86}$$

given a conjectured decision rule $\tilde{a}(\cdot)$.

However, note that this is an identical framework as the framework assuming θ is common knowledge if one defines $\tilde{\kappa} = \kappa\frac{\tau_\theta}{\tau_\theta+\tau_\mu}$. Given Proposition 11, outside the central pool the decision rule must be linear. All that remains to show is that the central pool is empty.

Suppose by contradiction, a non-empty central pool existed with pooling decision $a^* \leq 0$ and a pool determined by $\underline{v}, \bar{v}$. In this equilibrium, the player with preferences $\bar{v}$ must be indifferent between $\alpha\bar{v}$ and a^*. As the linear equilibrium can be extended to be an equilibrium over the entire real line, the player with preferences $\bar{v}$ prefers choosing $\alpha\bar{v}$ where $\phi(b, \alpha\bar{v}) \sim N(\frac{\tau_\theta}{\tau_\theta+\tau_\mu}\bar{v}, \frac{1}{\tau_\theta+\tau_\mu})$ to choosing $a = 0$ where $\phi(b, 0) \sim N(0, \frac{1}{\tau_\theta+\tau_\mu})$. Finally, note that choosing the pooling decision incurs a worse adaptation loss than $a = 0$ in the linear equilibrium (as $a^* \leq 0$), a worse conformity loss (as $a_1 = 0$ generates the ideal mean perceived preferences), and a worse variance as the equilibrium variance is higher in the pool. Finally, note if instead $a^* > 0$, one can utilize an identical argument with $\underline{v}$ to derive a contradiction. □

9.4.2 Learning About the Preferences of Others:

Let us now consider an alternative environment in which $\theta = 0$ is common knowledge, but there exists aggregate uncertainty over the average preferences of the population, μ. By induction, we can again drop any time dependence. Note that v is the only decision-relevant private information, implying a is a function of v alone. Moreover, a player's utility is:

$$-(a-v)^2 - \kappa \int \Big(\int \phi(b,a)(b-\mu)^2 db \Big) f(\mu \mid v) d\mu. \tag{87}$$

As in the proof of Proposition 3, I will denote by $\tau_{v_{\text{ind}}}$ the precision of v conditional on a realization of μ. I can now characterize the equilibria which satisfy D1.

Proposition 13 (**Linear Equilibria with State Uncertainty**)
The only equilibria that satisfy D1 off-path are the linear equilibria.

Proof of Proposition 13. Again, using a proof identical to Bernheim (1994), one can show that D1 implies central pooling. Outside the central pool, a player's preferences are fully revealed. Hence, the loss function outside the central pool is:

$$-(a-v)^2 - \kappa \int \big(v^{-1}(a) - \mu\big)^2 f(\mu \mid v) d\mu. \tag{88}$$

However a player's equilibrium perception of μ is a Gaussian distribution with a mean of $\frac{\tau_{v_{\text{ind}}}}{\tau_{v_{\text{ind}}}+\tau_\mu}v$ and a fixed variance. Given the quadratic-loss framework, the variance is additively separable from the decision rule and can be ignored when considering which decision to take outside the central pool. Let us recall the notation $\hat{v}(a)$ to denote the expected perceived preferences given decision a. Thus, the optimization reduces to choosing a a which maximizes,

$$-(a-v)^2 - \kappa\Big(\hat{v}(a) - \frac{\tau_{v_{\text{ind}}}}{\tau_{v_{\text{ind}}} + \tau_\mu} v\Big)^2. \tag{89}$$

One can differentiate to derive,

$$a - v + \kappa\Big(\hat{v}(a) - \frac{\tau_{v_{\text{ind}}}}{\tau_{v_{\text{ind}}} + \tau_\mu} v_1\Big)\hat{v}'(a) = 0. \tag{90}$$

Further, in equilibrium $\hat{v}(a) = v$, and thus,

$$a - v + \kappa\Big(v - \frac{\tau_{v_{\text{ind}}}}{\tau_{v_{\text{ind}}} + \tau_\mu} v\Big)v'(a) = 0. \tag{91}$$

However, this expression is identical to Equation (73) with a rescaling of κ. Thus the only differential equations that satisfy such a constraint are linear. Further, by an identical intuition to Proposition 11, the central pool must be empty in any equilibrium with revelation. These results imply any equilibrium with revelation that satisfies D1 is linear. □

9.4.3 Learning About Both the Preferences of Others and the Fundamental State:

This analysis is qualitatively different to the previous analysis. Now a_t is a function of two-dimensional private information: s_t, v_t. Thus, the equilibrium decision rule in general will be determined by a partial differential equation. Further, for any conjectured non-linear decision rule, $\phi(b, a_t)$ will be non-Gaussian. As one must integrate $\phi(b, a_t)$ to determine the conformity loss, one cannot solve (let alone, write down) the partial differential equation in closed form as was done for the differential equation in Proposition 11.

9.4.4 Unstable Confounded Learning Outcomes

Given the updating rule in Equation (35), there is a possibility that despite an equilibrium with revelation in every period that social learning fails. For instance, if $a_t = \lambda s_t + (1-\lambda)v_t$ in every period, the players' will never be able to disentangle θ and μ. The following proposition formalizes this intuition.

Lemma 5 *If there exists infinitely many periods of revelation, the variance matrix of the* *belief, Σ_t, either converges to the zero-matrix or Σ_t converges to*

$$\begin{pmatrix} c_1 & -\sqrt{c_1 c_2} \\ -\sqrt{c_1 c_2} & c_2 \end{pmatrix} \tag{92}$$

for $c_1, c_2 > 0$.

Proof of Lemma 5. Suppose that $\Sigma_t \not\to_p 0$, then Lemma 1 implies that $\theta(t) \not\to_p \theta$ and $\mu(t) \not\to_p \mu$ or neither. Further, for all t, a_t is a linear combination of s_t and v_t plus a constant. As a result, for all t a sufficient statistic for a_t is,

$$\lambda_t s_t + (1 - \lambda_t) v_t, \tag{93}$$

for some constant $\lambda_t \in (0, 1)$. If λ_t does not converge, then there exist two different convergent subsequences that converge to λ^1, λ^2, respectively. By the law of large numbers then the players learn $\lambda^1\theta + (1 - \lambda^1)\mu$ and $\lambda^2\theta + (1 - \lambda^2)\mu$. However, as $\lambda^1 \neq \lambda^2$ the players can separately learn θ and μ. Thus, because the beliefs do not converge, λ_t must converge.

If λ_t converges, then ϕ_t from Equation (35) necessarily converges to an interior value. Further, ρ_t is equal to the negative square root of $\tau_{2,t}^2/(\tau_{1,t}\tau_{3,t})$ as stated below:

$$\rho_t = -\sqrt{\frac{(\sum_{i=1}^t \sqrt{\phi_t(1 - \phi_t)})^2}{(\sum_{i=1}^t \phi_t)(\sum_{i=1}^t 1 - \phi_t)}}. \tag{94}$$

As ϕ_t converges, one can re-write ρ_t asymptotically as follows,

$$\rho_t = -\sqrt{\frac{t^2 \lim \phi_t(1 - \phi_t) + o(t^2)}{t^2 \lim \phi_t(1 - \phi_t) + o(t^2)}} \to -1. \tag{95}$$

Finally, the variance of θ and μ are each bounded and decreasing sequences, and thus both variances converge completing the proof. $\square$

I will call the above such beliefs "confounded learning outcomes." In this model, such beliefs are not asymptotically stable because after a small perturbation (such as decreasing the variance of θ) the beliefs will never return to this confounded learning outcome. Generally for such outcomes to be stable, there must exist multiple preference types who have opposing preferences (cf. Smith and Sørensen, 2000). By introducing γ_t, this ensures that if the signalling equilibrium involves revelation in each period then λ_t does not converge.