The Probability of Pluralistic Ignorance

Mauricio Fernández-Duque*

January 17, 2022

Abstract

I develop a theory of group interaction in which individuals who act sequentially are concerned with conforming to what they believe is the majority attitude. *Pluralistic ignorance* may arise, an outcome with incomplete learning in which individuals conform to a mistaken sense of the majority attitude and earning the majority's disapproval. The degree of uncertainty about the population distribution of attitudes affects what individuals learn about the group. A central finding is that the learning dynamics have a different impact on the probability of pluralistic ignorance in small and in large groups. I derive the maximum and minimum probabilities of pluralistic ignorance for groups of different sizes, as a function of the preferences for conformity and of uncertainty over attitudes. The theory provides hypotheses about underexplored questions regarding the prevalence of pluralistic ignorance.

Keywords: Pluralistic ignorance, group size, image concerns, expected judgments, social learning
JEL Codes: C72, D83, D90

*CIDE. Carretera México-Toluca 3655, Col. Lomas de Santa Fe, Del. Cuajimalpa, 01210, Mexico City, Mexico. E-mail: mauricio.fernandez@cide.edu. I would like to thank Samuel Asher, Alberto Alesina, Nava Ashraf, Robert Bates, Iris Bohnet, Filipe Campante, Georgy Egorov, Paul Feldman, Jeffry Frieden, Matt Gentzkow, Benjamin Golub, Michael Hiscox, Toren Iversen, Antonio Jiménez, Horacio Larreguy, Emir Kamenica, Navin Kartik, Scott Kominers, José Luis Montiel Olea, Stephen Morris, Carlo Prato, Pia Raffler, Matthew Rabin, Kenneth Shepsle, Joel Sobel, Levent Ulku, and seminar participants at Harvard University, ITAM, CIDE, and the EEA-ESEM 2018 meeting.

When members of a group act according to what they think others want, they may end up doing what nobody wants. In a classic paper, O'Gorman (1975) shows that the majority of whites in the U.S. in 1968 did not favor segregation. However, they overestimated support for it, with about half believing that the majority of whites *did* favor segregation. Furthermore, overestimating support for segregation made them more willing to support segregationist housing policies. O'Gorman was studying *pluralistic ignorance*, a situation in which "a majority of group members privately reject a norm, but incorrectly assume that most others accept it, and therefore go along with it" (Katz and Allport, 1931). Whites' public opinion regarding segregation is an example of pluralistic ignorance at a large scale. Pluralistic ignorance is also found at small scales. A racist remark may go uncontested in a conversation, since individuals may incorrectly believe that others approve of it. But not contesting the remark may strengthen its perceived approval (e.g. Kuran, 1997).

Pluralistic ignorance has been documented at large and small scales (e.g. Smerdon, Offerman, and Gneezy, 2020, Bursztyn, González, and Yanagizawa-Drott, 2020) and has been argued to lead to persistence in social norms and to rapid social change (Bicchieri, 2005, Bursztyn, Egorov, and Fiorin, 2017, Sunstein, 2019), but there is a lack of work on how likely it is to arise. This paper studies the probability of pluralistic ignorance when social learning shapes image concerns. In other models of conformity, the preferred type or action is commonly known (e.g. Bernheim, 1994, Ali and Bénabou, 2020). In this model, individuals aim to conform to the majority attitude, which they do not know, but infer by combining their private information with the history of play, based on a common prior.[1] Pluralistic ignorance arises when individuals conform to what they misperceive as the majority attitude.

Many basic theoretical questions about pluralistic ignorance are open. Under what conditions does pluralistic ignorance emerge in small groups and in large groups, and what is the probability with which it arises? How does the probability of pluralistic ignorance change as the environment changes? Can we predict the probability of pluralistic ignorance by the prevalence of certain actions, or the other way around? Can we use data on pluralistic ignorance to make inferences

[1] I use the term "attitudes" in this section and the next, and "types" for the rest of the paper.

about whether agents are acting rationally? The main theoretical contribution of the model is to characterize the probability with which pluralistic ignorance arises, and to show that the forces driving pluralistic ignorance differ between small and large groups. The predictions yielded by the model are meant to guide empirical research towards unaddressed questions about the probability of pluralistic ignorance.

The key features of the model are as follows. The basic setting I have in mind is a conversation about attitudes towards a socially charged topic, such as segregation. Individuals have private attitudes, which are informative about the population from which their group is drawn. Individuals wish to conform to the majority attitude in their group, and can observe others' actions since decisions are sequential. However, social learning is limited since the desire to conform may lead individuals to not reveal their attitude. This incomplete learning gives rise to pluralistic ignorance. The model considers finite groups of any size, and the probability of pluralistic ignorance turns out to depend differently on initial conditions in small and large groups. The main contribution of the model relative to previous work is that, by including the key ingredients of private attitudes, uncertainty about the population, sequential decisions, a desire to conform, and finite groups of any size, the model permits a definition of pluralistic ignorance that captures its important features, and permits a tractable analysis of its probability. Crucially, the model allows me to show how this probability depends on group size.

In the model, individuals interact in a group by sequentially expressing an attitude. Each individual has a privately observed *true attitude*, but chooses which attitude to express. For example, individuals may have a pro or anti-segregation attitude, but may express either attitude. Others in the group judge those who have expressed an attitude. Alice's judgment of Bob is increasing in Alice's belief that Bob's true attitude is equal to her own true attitude. I focus on the subset of Perfect Bayesian Equilibria in which individuals express their true attitude if they expect most in the group share their true attitude. This restriction imposes a natural monotonicity on strategies: an individual is more likely to express an attitude the more likely others will judge him positively for doing so. Bob and Alice are both fully Bayesian, have correct common priors, and take into account all available information. Public information and social dynamics may nevertheless lead Bob to have mistaken beliefs over how Alice will judge him, and to act in a way that reinforces Alice's mistaken beliefs over how Bob will judge her.

I have found it useful to break down the definition of pluralistic ignorance into three features. A contribution of this paper is to provide a formal definition of pluralistic ignorance which captures these three features, and I will study when they appear in equilibrium with a positive probability. First, most group members do not express their true attitude—many whites expressed support for segregation even though they did not really support it. Second, group members overestimate how many others express their true attitude—whites believed that those who expressed support for segregation did really support it. Third, overestimating how many express their true attitude makes individuals conform by not expressing their own true attitude—whites were more willing to express support for segregationist housing policies when they overestimated others' support.

Most of the time pluralistic ignorance will not arise, given that individuals are Bayesian. Nevertheless, there may be a non-trivial probability of pluralistic ignorance, or more precisely, a non-trivial probability that group members are assigned true attitudes that lead to pluralistic ignorance in equilibrium. In Section 3 I show that, for extreme enough parameter values, the probability of pluralistic ignorance is arbitrarily close to $1/2$.

To understand when pluralistic ignorance arises, we need to understand how an individual forms his expectation of how he will be judged for expressing an attitude. He must consider what he knows about group members' true attitudes. This will depend on common priors, his private information, what others have revealed about their true attitude if they have already expressed an attitude, and how many group members there are. Someone writing a Twitter post will take into account that anyone can see the post, and will use others' past tweets to form beliefs about the potentially large audience. An individual in a private conversation will use what others in the conversation have said to form beliefs about them. It will turn out that pluralistic ignorance arises under different circumstances depending on whether the group is *large* (as with Twitter posts) or *small* (as during a private conversation).

In addition to the size of the group, another key dimension of the environment is *second-order uncertainty* over group attitudes, i.e. uncertainty over the composition of the population that true attitudes come from. For simplicity, I posit that true attitudes are drawn from one of two populations with different majority attitudes, and group members commonly know the probability with which they were drawn from each of the populations. Second-order uncertainty affects which attitudes are expressed. When there is *low* second-order uncertainty, all individuals strongly believe they are drawn from a specific population, so all expect to be mostly judged according to the majority attitude in that population. They therefore all express the majority attitude of that population. When there is *high* second-order uncertainty, first movers believe many share their true attitudes. The reason is that one's own true attitude provides a signal of the population from which they are drawn, a signal that is very informative with high second-order uncertainty. First movers express their true attitude, believing most will judge them positively, and in so doing reveal information about the majority attitude in the group. Once enough information has been revealed, individuals ignore their own signal and herd on the perceived majority attitude.

Note that although individuals' true attitudes do not change in the model, the different ranges of second-order uncertainty could be understood as capturing either a situation of stability in true attitudes (when there is low second-order uncertainty), or one in which true attitudes were in flux, which has led to uncertainty (when there is high second-order uncertainty). The Jim Crow era in the South of the U.S. featured low second-order uncertainty for whites regarding segregation, since there was a strongly perceived pro-segregation consensus. The Civil Rights movement and its aftermath were a period of flux regarding segregation attitudes in the U.S., featuring high second-order uncertainty.

The main idea of the paper is that second-order uncertainty and group size together are key to understanding the probability of pluralistic ignorance. The main result characterizes the intervals of second-order uncertainty in which the probability of pluralistic ignorance is highest or lowest.[2]

[2]A finer characterization of the exact probabilities of pluralistic ignorance is difficult due to some "discrete" features of the model—finite group size and binary attitudes. These discrete features, in turn, are necessary for the key behavior of the model—pluralistic ignorance that arises in equilibrium in groups of any size.

The probabilities are lowest among small groups when second-order uncertainty is high, lowest among large groups when second-order uncertainty is low, and highest among small and among large groups when second-order uncertainty is neither high nor low. Group size matters because it affects the uncertainty individuals have about the realized group. With low second-order uncertainty, the uncertainty over the realized group goes to zero as group size increases (by the law of large numbers). With high second-order uncertainty, the first movers who reveal their type are a large fraction of small groups, so they provide a lot of information about the group composition.

Although herding dynamics are an important part of my model, pluralistic ignorance is conceptually different from incomplete learning in standard models of herding or informational cascades (Bikhchandani, Hirshleifer, and Welch, 1992, Banerjee, 1992, Smith and Sørensen, 2000, Frick, Iijima, and Ishii, 2020—see Golub and Sadler, 2016 for a review). In those models, actors are not trying to express an attitude to match the majority of true attitudes in the group, they are trying to match a state of the world which is not determined by others' true attitudes. Therefore, the definition of pluralistic ignorance does not properly apply: individuals in those models choose the wrong action due to a mistaken sense of what is best for them regardless of others' true attitudes, not because they conform to a majority they misperceive.

Pluralistic ignorance has often been modeled as the inefficient outcome of a coordination game (e.g. Chwe, 1999, Kuran, 1989, Chwe, 1999, Bicchieri, 2005, Centola, Willer, and Macy, 2005, Smerdon et al., 2020). The motivation to coordinate, however, does not give rise to a misunderstanding *per se*: actors may be stuck in an inefficient equilibrium, but know they are stuck there without any misunderstanding of others' preferences. These models feature multiplicity of equilibria which makes it hard to predict which outcome will arise, although recent work has tried to get around this problem by allowing preferences to change exogenously (Smerdon et al., 2020, Andreoni, Nikiforakis, and Siegenthaler, 2020, Duffy and Lafky, 2020). Furthermore, since actors are motivated to do what others do, the equilibria that arise generally do not vary with group size.

The forces in this paper do not arise in models with reputational concerns in which individuals either try to influence others or in which they try to conform, since those models do not feature aggregate uncertainty about how others judge an action or announcement. When mistaken beliefs arise in the models where individuals try to influence others, it is due to some individuals misleading others to act in ways that benefits them (Morris, 2001 and Gentzkow and Shapiro, 2006, Che, Dessein, and Kartik, 2013). In my model, aggregate uncertainty may lead individuals to act in a way which benefits *nobody*, by trying to conform to a mistaken view of the majority. Models in which individuals try to conform have either featured correct equilibrium beliefs and no aggregate uncertainty (e.g. Prendergast, 1993, Bernheim, 1994, Bénabou and Tirole, 2006, Levy, 2007, Visser and Swank, 2007), or pluralistic ignorance is assumed *ex ante* instead of derived endogenously (e.g. Benabou and Tirole, 2012, Bursztyn et al., 2020). In these second types of models, and in line with the absence of aggregate uncertainty, the actors who care about conforming do not take turns choosing actions that may reveal their private information, and as such herding dynamics and their potential for misinformation do not arise.

Empirical work has documented many instances of pluralistic ignorance in large and small groups, providing evidence that the phenomenon exists and can have a large impact on behavior (e.g. Prentice and Miller, 1993, Shamir and Shamir, 2000, Wenzel, 2005, Mildenberger and Tingley, 2017). For instance, Bursztyn et al. (2020) document pluralistic ignorance among Saudi men regarding attitudes towards female labor-force participation, and show that correcting the misinformation increases sign-ups to costly job-matching applications by 57% from a baseline of 21% among the misinformed. However, to my knowledge there is almost no work that studies the pervasiveness of pluralistic ignorance, or the follow-up questions of which forces affect its pervasiveness, and how those forces differ between small and large groups. By deriving testable comparisons on the probability with which pluralistic ignorance arises across settings, I establish theoretical guidance for further empirical study of this important phenomenon.

1 A Minimal Example and a Roadmap

Here I present a minimal example to illustrate the main results.[3] I use the minimal example to provide a roadmap for the rest of the paper.

Suppose there are two individuals, $i \in \{A, B\}$, who are having a conversation about segregation. At the beginning of the conversation, individuals privately observe whether they are actually pro-segregation ($\theta_i = H$) or anti-segregation ($\theta_i = L$)—this is their true attitude. The conversation is highly stylized: first, Alice expresses a pro-segregation attitude ($a_A = H$) or an anti-segregation attitude ($a_A = L$), and then Bob expresses a pro-segregation attitude ($a_B = H$) or an anti-segregation attitude ($a_B = L$).

Alice gets one util from expressing her true attitude, and zero utils otherwise. After she expresses an attitude, Bob passes a positive or negative judgment over her expressed attitude. Bob's judgment is positive if Alice's type is equal to his own type with a probability of at least $\overline{J} \in (0.5, 0.545)$. Alice puts weight $\beta > 10/3$ on her expectation of a positive judgment by Bob, which I call her *expected judgment*. She cares about being judged positively (perhaps due to reputational concerns), and trades that off against expressing her true attitude. Thus, Alice's expected utility is:

$$\mathbb{1}\{a_A = \theta_A\} + \beta P\big(P(\theta_A = \theta \mid a_A, \theta_B = \theta) > \overline{J} \mid \theta_A\big),$$

Alice's expected judgment is the probability she places on Bob believing she is of the same type with at least probability $\overline{J}$. Notice the importance of second-order beliefs when forming expected judgments.

Bob's expected utility is similar to Alice's, but before forming his expected judgment, Bob observes the expressed attitude of Alice:

$$\mathbb{1}\{a_B = \theta_B\} + \beta P\big(P(\theta_B = \theta \mid a_A, a_B, \theta_A = \theta) \geq \overline{J} \mid \theta_B, a_A\big).$$

How does this conversation play out? It depends on the information the individuals have of

[3] The full details of the analysis are in Online Appendix OA.

each other at the outset. In particular, it depends on second-order uncertainty: the uncertainty over the *ex ante* probability with which a group member is pro-segregation. This section will focus on two extreme examples of this uncertainty.

A lack of second-order uncertainty may lead Alice and Bob to express a pro-segregation attitude no matter their true attitude. Suppose they both believe they were drawn from a population with 65% pro-segregationists. There is no second-order uncertainty since they commonly know the probability with which a group member is pro-segregation. Since Alice believes Bob is most likely pro-segregation, she believes she is very likely to be judged positively if she expresses a pro-segregation view, and negatively if she expresses an anti-segregation view. Alice then expresses a pro-segregation view whatever her true attitude, which leaves Bob with the same information Alice had when she made her decision. Therefore, Bob expresses a pro-segregation view whatever his true attitude.

High second-order uncertainty leads Alice to express her true attitude, and Bob to express whichever attitude Alice expressed, whatever his true attitude. Suppose there is a 50% chance that the population from which their attitudes are drawn is 65% pro-segregation, and a 50% chance that it is 65% anti-segregation. To form beliefs about the population from which they were drawn, Alice uses her private information—her true attitude. As a Bayesian, she puts more weight on the population having a majority type that equals her type. But then she believes most judges are of her type, and prefers to express her true attitude. Since Alice expressed her true attitude, and in equilibrium Bob knows this, Bob knows how Alice will judge him. He therefore expresses Alice's true attitude no matter his own true attitude, knowing that Alice will judge him positively since she will believe he is most likely of her type.

Pluralistic ignorance is defined for a specific realization of true attitudes. For that realization, in equilibrium most individuals express an attitude that differs from their true attitude, and individuals reach the end of the game expecting most expressed their true attitude. This would require both individuals to be anti-segregation and express a pro-segregation attitude, or vice versa.

With high second-order uncertainty, the probability of pluralistic ignorance is 0: since Alice always expresses her true attitude, in all realizations of true attitudes at least half the individuals express their true attitude. With no second-order uncertainty, the probability of pluralistic ignorance is positive. When both individuals are anti-segregation (an event with 12% probability),[4] both will express a pro-segregation attitude. Moreover, since neither individual revealed information, at the end of the game they will continue to believe the other individual is most likely pro-segregation, and therefore expressed his or her true attitude.

Roadmap. After setting up the general model in Section 2, in which group size and second-order uncertainty are parameters, Section 3 shows that two equilibrium dynamics appear as a function of second-order uncertainty, mirroring those of the minimal example. When the second-order uncertainty is low, no individuals reveal their true attitude—they all express whichever true attitude is more widespread in the population from which they all strongly believe they are drawn (Proposition 1, Section 3.1). When second-order uncertainty is high, the first to express an attitude express their true attitude. After a large enough run of individuals revealing the same true attitude, the rest all herd on expressing that same attitude (Proposition 2, Section 3.2). In the minimal example, this run was of size 1. Propositions 1 and 2 also derive the probability of pluralistic ignorance with high and low second-order uncertainty in the general model. Proving Proposition 2 turns out to be the most complicated part of the analysis, since individuals sometimes randomize how much they reveal about themselves, and beliefs over judgments depend on how much information others have revealed about themselves.

In Section 3.3 I present Proposition 3, the main result of the paper: the probability of pluralistic ignorance is lowest in small groups when second-order uncertainty is high, lower in large groups when second-order uncertainty is low, and highest in small and large groups when second-order uncertainty takes on intermediate values. I further present some comparative statics, provide a way to bound the size of small and large groups, and show the non-monotonic relationship between the probability of pluralistic ignorance and the probability of choosing an action.

Section 4 discusses the relation to existing evidence.

[4] The probability is $0.35 \times 0.35 = 0.1225$, since they are drawn with certainty from a population with 35% anti-segregationists.

2 Setup

In the model there is a group of individuals who act sequentially with the rest judging them. Actors are motivated to act in a way that signals a type that is judged positively by the judges. The optimal action therefore depends on the information actors and judges have about each others' type. Worked-out numerical examples of the model can be found in Section 1, and Appendices OA and OB.

Individuals and periods. There is a group with $I \geq 2$ individuals and $2 + I \times 2$ periods, with individuals labeled $i, j, k, l \in \{1, 2, ..., I\} \equiv I$, and set of periods $\{-1, 0, 1, 1.5, 2, 2.5, ..., I, I.5\}$. I abuse notation by using I to denote both the set and number of individuals, as it will not lead to ambiguity.

Information structure. Individuals are endowed with a privately observed *type* $\theta_i \in \{H, L\}$.

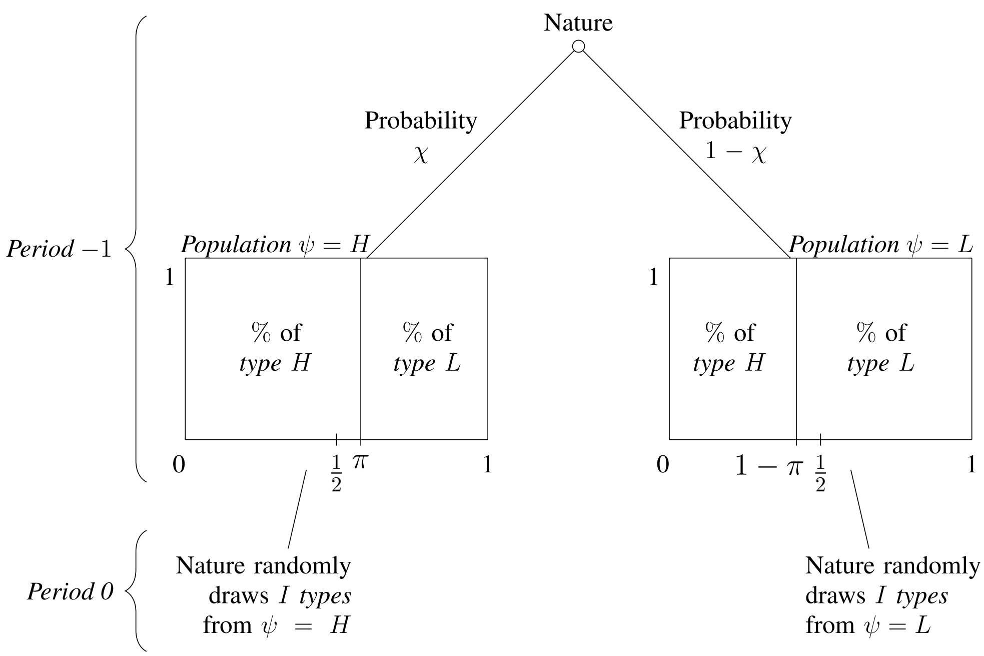

Figure I: Assignment Of Realization Of Types, An I-Tuple $(\theta_1, \ldots, \theta_I)$. Population ψ and others' types are not observed; the context χ and the precision π are common knowledge.

Figure I is a graphical representation of how individuals' types are drawn, which results in a *realization of types* $(\theta_1, \ldots, \theta_I)$. Nature selects a population $\psi \in \{H, L\}$ from which individuals are drawn, selecting population $\psi = H$ with the commonly known probability χ. I will refer to

χ as the *context*. Individuals are randomly drawn from the selected population, and are randomly indexed from 1 to I. Conditional on the population ψ being equal to $\theta \in \{H, L\}$, an individual's type is equal to θ with probability $\pi \equiv P(\theta_i = \theta \mid \psi = \theta) \in [1/2, 1)$. Population $\psi = \theta$ is more likely to produce groups with majority type θ. Since π indicates how likely Nature is to draw the majority type in a given population, I will refer to it as the *precision* of the population. Although the population ψ is not observed and an individual's type θ_i is private information, both the context χ and the precision π are commonly known. Notice that types are informative about the population.

I focus on the range $\chi \geq 1/2$. Given the symmetry of the setup, this is without loss of generality.

Actors. Once the group is assigned, individuals take turns choosing actions and judging each others' actions. In period i, individual i chooses *action* $a_i \in \{H, L\}$ after observing the history of play $h_i = (a_1, a_2, ..., a_{i-1})$, with $\mathcal{H}_i$ the set of these histories. As a shorthand, whenever actor i chooses the action that corresponds to his type ($a_i = \theta_i$), I will say he *chooses his type*. Figure II denotes the timeline.

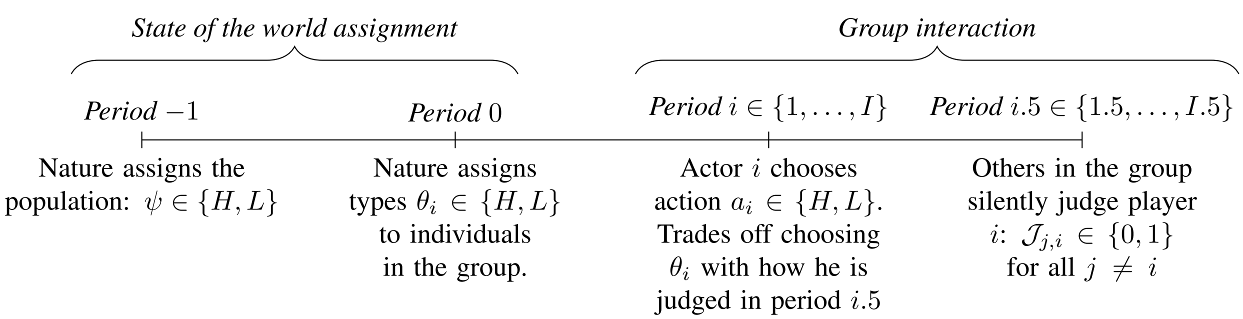

Figure II: Timeline

In period $i.5$, which comes after i acts but before $i + 1$ does, all individuals other than i judge i based on his decision. Individual i trades off choosing his type with acting according to how he expects to be best judged by others. Let $\mathcal{J}_{j,i}(a_i) \in \{0, 1\}$ be j's judgment of i. As I will elaborate shortly, j judges whether i's type matches her own type. Individual i's payoff is affected by the average of expected judgments as follows:

$$\mathbb{E}u(a_i; \theta_i, h_i, I) = \mathbb{1}\{a_i = \theta_i\} + \frac{\beta}{I-1}\sum_{j \neq i} \mathbb{E}\Big(\mathcal{J}_{j,i}(a_i) \mid h_i, \theta_i\Big) \tag{1}$$

with $\mathbb{1}\{\cdot\}$ the indicator function. I refer to the first summand as the *material payoff*, to an individuals' expectation over average judgments as his *expected judgment*, and to β as the *weight on the expected judgment*.

I assume $\beta > 1$—individuals would always choose their type if $\beta \in [0, 1]$. Concerns over the expected judgment can be justified as a reduced-form reputational concern, such as wanting to be judged positively by most individuals in the group, or equivalently, by as many as possible. Note that I will generally use i for the individual acting in the current period, or the *actor* (he), and will use j for a *judge* of i (she).

Judges. When judge j judges actor i, judge j must decide whether to *judge i positively* ($\mathcal{J}_{j,i} = 1$) or *negatively* ($\mathcal{J}_{j,i} = 0$). Judge j sets $\mathcal{J}_{j,i} = 1$ if j believes the probability i's type matches her own is more than the *judgment threshold* $\overline{J}$ (that is, $P(\theta_i = \theta \mid h_i, a_i, \theta_j = \theta) > \overline{J}$). If j believes the probability is less than $\overline{J}$ (that is, $P(\theta_i = \theta \mid h_i, a_i, \theta_j = \theta) < \overline{J}$), then $\mathcal{J}_{j,i} = 0$. In the intermediate case of $P(\theta_i = \theta \mid h_i, a_i, \theta_j = \theta) = \overline{J}$ judge j is indifferent in her judgment.

Judgments are not observed by others, perhaps because judges do not want to or cannot make them immediately public. An actor does not observe how he is judged by the end of the game.

A natural way to motivate the judgment function $\mathcal{J}_{ji}$ is that a positive judgment is a decision to engage positively with the actor once the game has ended, for example by paying attention to him, following or liking him on social media, or just thinking highly of him. Judges prefer a positive judgment for those who share their type, and a negative judgment for the rest. Judges are biased towards judging others negatively; perhaps they are selective in who they like because of scarcity in their attention or social capacity. The judges' utility function can then be written:

$$\mathbb{1}\{\mathcal{J}_{ji} = 0\} + \delta[\mathcal{J}_{ji}\mathbb{1}\{\theta_i = \theta_j\} + (1 - \mathcal{J}_{ji})\mathbb{1}\{\theta_i \neq \theta_j\}]$$

where $\delta > 0$. This function yields a threshold rule such that judge j judges i positively if and only if $P(\theta_i = x \mid h_i, a_i, \theta_j = x) > (\delta + 1)/2\delta \equiv \overline{J}$. The assumption that judges are biased towards judging others negatively implies that $\overline{J} > 1/2$. I further assume that $\overline{J} < \pi^2 + (1 - \pi)^2$, which implies that judges are not too negatively biased in their judgments.[5]

To form beliefs about the actor's type, the judge uses the information she has right after the actor's action, considering what the actor knew when making a decision. In ongoing work (Fernández-Duque, 2021a), I consider a more flexible timing of judgments.

Comments on setup. A binary choice provides the simplest setting for individuals to reveal their types through their actions, or to remain "silent" in the sense of not revealing their type. This sequential decision-making captures in a stylized way the dynamics of expressing attitudes—individuals do not all express their attitude at once, and past actions inform later ones. Note that the more general terms "types" and "actions" replace the terms "true attitudes" and "expressed attitudes" I used in earlier sections, while the "context" captures what I referred to as "second-order uncertainty".

Types impact utility through two channels. Since they are signals about the realization of types, they affect judges' judgments and actors' expected judgment. In addition, they directly affect actors' utility through the first summand in the utility function (1). The first channel can be uncorrelated from the second by providing individuals additional private signals of the population. The approach I am taking is a useful simplification. Moreover, a propensity to believe others are likely of one's type is consistent with psychological evidence.[6]

The uncertainty over judgments is the heart of the dynamics of the model. If there were no uncertainty over the distribution of a group's types, actor i would not face any uncertainty over

[5]If judges are too negatively biased, actors will be judged negatively even if they fully conform. This leads to mixed strategies in groups of arbitrarily large size, which makes the analysis much less tractable. Intractability due to mixed strategies also arises if instead of discrete judgments, judgments were linear in the probability $P(\theta_i = \theta \mid h_i, a_i, \theta_j = \theta)$ as in Bursztyn et al. (2017) and Bursztyn et al. (2020). Mixed strategies would arise because actors who would otherwise be willing to conform to a perceived majority type θ may not be judged sufficiently positively by judges of type θ.

[6]Ross, Greene, and House (1977) coined the term "false consensus effect" to refer to the finding that individuals overestimate the proportion of others' types that matches their own type. As Dawes (1989) pointed out, the effect may be driven by rational updating—using one's own type to infer others' attitudes. This argument is echoed by our setup. More recent experimental work has studied the rationality of this effect (Engelmann and Strobel, 2000, 2012).

how he'll be judged. The result would be a simplified model of conformity as in Bernheim (1994), where all want to signal what they commonly know is the majority type, but where actions are sequential.[7] In fact, this is the special case of my model in which the context is equal to 1 and the group is arbitrarily large. Individuals in the special case are arbitrarily confident that the majority type is H by the law of large numbers.

A judge's judgment compares two types: her own, which she knows perfectly, and the actor's. To form beliefs about the actor's type, the judge uses the information she has right after the actor's action, considering what the actor knew when making a decision.

2.1 Strategies and Equilibrium

Denote by $G \equiv (\chi, \pi, \beta, I)$ the game defined above with context χ, precision π, weight on expected judgment β and group size I. A strategy of i is $\sigma_i : \mathcal{H}_i \times \{H, L\} \to [0, 1]$, which maps a history h_i and a type θ_i to a probability of choosing H. The *strategy prescription* σ_i evaluated at h_i, or $\sigma_i(h_i)$, indicates the probability with which each type of i chooses H given history h_i—in other words, $\sigma_i(h_i)$ is a type-dependent strategy.

I say that actor i *pools on* $a \in \{H, L\}$ at history h_i if $\sigma_i(h_i)$ prescribes i to choose a no matter his type. This is what I have meant so far when I have said that an actor "conforms". I say actor i *separates* at history h_i if $\sigma_i(h_i)$ prescribes i of either type to choose his type. I say actor i *semi-pools on* a at history h_i if $\sigma_i(h_i)$ prescribes i of type a to choose a for sure, and for i of the type that is not a to randomize. If it does not lead to ambiguity, I will not mention the history when talking about a separating, a semi-pooling or a pooling strategy prescription.

I will restrict attention to a subset of Perfect Bayesian Equilibria that are natural in this setting. I use the intuitive criterion to refine out-of-equilibrium beliefs (Cho and Kreps, 1987). The intuitive criterion posits that, if an action is chosen that no type would choose on the equilibrium path of

[7] The literature on conformity and social image concerns typically assumes common knowledge over the type that individuals wish to signal (e.g. Bernheim, 1994, Morris, 2001, Bénabou and Tirole, 2006, Sliwka, 2007, Ellingsen, Johannesson, et al., 2008, DellaVigna, List, and Malmendier, 2012, Ali and Bénabou, 2020, and see Bursztyn and Jensen, 2017 for a review).

play, individuals will infer that the deviation came from a type who would deviate for *some* out-of-equilibrium beliefs. I use the intuitive criterion to define beliefs corresponding to action $-a$ whenever an actor pools on a.

Definition 1. *A strategy prescription $\sigma_i(h_i)$ is **monotonic** if it prescribes that type θ chooses θ at history h_i if he expects at least half of the judges are of type θ ($\mathbb{E}_{j \neq i} P(\theta_j = \theta \mid h_i, \theta_i = \theta) \geq 1/2$).*

*A **Monotonic PBE (MPBE)** is a PBE where actors follow monotonic strategy prescriptions at all histories. A **MPBE strategy** is a strategy followed in a MPBE.*

Monotonicity implies that actors maximize their material payoff as long as enough judges will judge the actor positively for doing so. This condition imposes consistency in PBE strategies: if actor i chooses $a \in \{H, L\}$ when following a strategy prescription that type a chooses a, then actor $i + 1$ follows a strategy prescription that type a chooses a.[8] Moreover, the condition avoids unnatural equilibrium strategies in which actors choose the opposite of their type in order to signal their type, such as a strategy where type H chooses L and type L chooses H. The results below will consider the common properties of all MPBE.

Actor i *uniquely pools* on a when he has a unique MPBE strategy prescription to pool on a at history h_i. Note that although an actor may uniquely pool on a, pooling on H and pooling on L may both be strategy prescriptions of Perfect Bayesian Equilibria.

2.2 Defining Pluralistic Ignorance

In this section I define pluralistic ignorance and the probability of pluralistic ignorance.

Definition 2. *There is **pluralistic ignorance** for realization of types $\overline{\theta} \equiv (\theta_1, \theta_2, ...\theta_I)$ in game G with strategy profile σ if and only if, by the end of the game,*

[8] This type of consistency would not hold if I imposed the following weaker assumption on strategies: at every history h_i, $P(a_i = H \mid h_i, \theta_i = H) \geq P(a_i = H \mid h_i, \theta_i = L)$, which means that action H is a signal that the type is H. With this weaker assumption, I cannot use the intuitive criterion to pin down out-of-equilibrium beliefs when actors pool, so an actor may pool on H and on L in distinct equilibria. There would then be multiple equilibria in which actors alternate which action they pool on, which affects whether there is pluralistic ignorance.

1. *The average probability that actors did not choose their type is greater than* $1/2$*:*

$$\sum_i \frac{P(\theta_i \neq a_i \mid \overline{\theta}, t = I.5)}{I} > 1/2.$$

2. *Individuals expect the average probability that actors did not choose their type is less than* $1/2$*:*

$$\mathbb{E}_i \left(\sum_{k \neq i} \frac{P(\theta_k \neq a_k \mid \theta_i, t = I.5)}{I - 1} \middle| \overline{\theta} \right) < 1/2 \quad \textit{for all } i.$$

The ***probability of pluralistic ignorance*** *of game* G *with strategy profile* σ *(conditional on a subset of realizations of types* $\widehat{\Theta} \subseteq \{(\theta_1, \ldots, \theta_I) \mid \theta_i \in \{H, L\}\}$*) is the proportion of realization of types (conditional on* $\widehat{\Theta}$*) that result in pluralistic ignorance. The* ***MPBE probabilities of pluralistic ignorance*** *of game* G *(conditional on* $\widehat{\Theta}$*) are the probabilities of pluralistic ignorance (conditional on* $\widehat{\Theta}$*) for some MPBE of game* G*.*

A bit loosely, if there is pluralistic ignorance, most actors in the group choose the opposite of their type, but they expect most chose their type. Pluralistic ignorance among whites in our running example would be a situation in which (a) most expressed support for segregation although they actually opposed it, and (b) they expect most expressed support for segregation and actually supported it. Alternatively, it could be a situation in which (a) most expressed opposition to segregation although they actually supported it, and (b) they expect most expressed opposition to segregation and actually opposed it. Substituting the strict inequalities for weak inequalities in the definition leads to similar results. Note that, by the definition, the misunderstanding raised by pluralistic ignorance persists until the end of the game.

Although the formal definition of pluralistic ignorance is novel, it captures the essential features of the original use of the term, quoted in the opening paragraph of the paper: most actors act in a way that they privately reject, but do so because they incorrectly assume that most others do not privately reject it and therefore conform to the misperceived majority type. As I mentioned in the introduction, I break down the definition into three features: most actors do not choose their type (part 1 of Definition 2), that actors overestimate how many choose their type (part 2 of Definition 2), and that overestimating how many chose their type makes actors conform. This third feature is not explicitly stated in Definition 2, but is implicit in the model since what motivates actors to not choose their type is the expected judgment term.[9] Note that actors would not act in a way that they privately reject without a large enough weight on expected judgments β.

I will sometimes condition the MPBE probability of pluralistic ignorance on a subset of the realizations of types such as those where all actors have a unique or a pure strategy prescription.

3 Results

This section presents the results of the model. I will show that there are two dynamics that can occur in equilibrium, and which one occurs depends on whether the context χ falls above or below a cutoff defined as follows:

Definition 3. *The* ***cutoff*** $\widehat{\chi}$ *for the game* G *is the minimum context such that actor 1 uniquely pools on* H.

In Appendix A I show that the cutoff $\widehat{\chi}$ is equal to:

$$\frac{\pi(\beta(2\pi-1)+1)}{(1-\pi)(\beta(2\pi-1)-1)+\pi(\beta(2\pi-1)+1)} \in (\pi, 1) \tag{2}$$

The cutoff $\widehat{\chi}$ is less than 1 if and only if $\pi > (\beta+1)/2\beta$, an assumption I will sustain throughout, since it is necessary and sufficient for the two equilibrium dynamics to arise for some contexts. The assumption implies that an actor would uniquely pool on θ if all he knew about the group is that it was drawn from population θ.

3.1 Actors Pool on H With Contexts Close To 1

I begin by stating the result.

[9]Since actors do not choose their type only in order to satisfy expected judgments, it turns out that the first condition of Definition 2 can be achieved in equilibrium only if the second condition is achieved.

Proposition 1. *Consider a game G with context $\chi \geq 1/2$ and $\pi > (\beta+1)/2\beta$.*

For all contexts above the cutoff $\widehat{\chi}$, all actors pool on H on the equilibrium path of play of the unique MPBE. The MPBE probability of pluralistic ignorance is the probability that most in the group are of type L. It is positive, decreases with the context χ, and tends to $1-\chi$ as the group size increases.

When the context is above the cutoff $\widehat{\chi}$, the first actor will expect more than half of the judges are of type H, independent of his type. He expects there are enough H-types that he will uniquely pool on H out of expected judgment concerns. Since actor 2 does not learn anything from actor 1's action, he will have the same incentives as actor 1 and will therefore also uniquely pool on H. So will all other actors.

Actors learn much about judges' types from the context, but nothing from others' actions. Pluralistic ignorance arises when individuals are mostly L-types, given that the context led them to expect they would be mostly H-types.

When all pool on H, the probability of pluralistic ignorance increases the more likely it is that most in the group are L-types. In contexts above the cutoff $\widehat{\chi}$, the lowest probability of pluralistic ignorance is induced by the context $\chi = 1$ and tends to 0 as group size grows, and the largest probability is induced by the context equal to the cutoff $\widehat{\chi}$, and tends to $1-\widehat{\chi}$ as group size grows. This result uses the law of large numbers. As the group size grows, the majority type in the group is increasingly determined by the population. For arbitrarily large groups then, the probability of pluralistic ignorance is arbitrarily close to the probability that the population ψ is L.

Although the probability of pluralistic ignorance can be made arbitrarily close to 0 in large groups, in small groups the probability is positive. No matter how much members of a small group know about the population, there is always a positive probability that the majority type in the group differs from that of the population.

3.2 Actors Herd with Contexts Close to $1/2$

The probability of pluralistic ignorance differs when the context is below the cutoff $\widehat{\chi}$ from when the context is above the cutoff, since in the former case actors herd once they have learned enough about the group from others' actions, whereas in the latter case they do not learn from others' actions. Understanding the difference in dynamics, and the corresponding difference in the probability of pluralistic ignorance, will further allow us to understand how the probability of pluralistic ignorance varies with the context and the group size.

Following Ali and Kartik (2012), I begin by defining the *lead* for action H at period i,

$$\Delta(h_i) \equiv \sum_{k=1}^{i-1} \mathbb{1}\{a_k = H\} - \mathbb{1}\{a_k = L\}.$$

This summary statistic keeps a tally of how many individuals have chosen action H at period i, and subtracts how many have chosen action L. I will refer to $-\Delta(h_i)$ as the lead for action L.

Definition 4. *A* ***herding dynamic*** *in G is a strategy profile in which actor 1 separates, and actor i separates if the lead for action H is between some thresholds: $\underline{N}(L, h_i) < \Delta(h_i) < \underline{N}(H, h_i)$. All actors from i onwards pool on θ if the lead for action θ is greater or equal to some threshold $\overline{N}(\theta, h_i)$, with $|\overline{N}(\theta, h_i)| \geq |\underline{N}(\theta, h_i)|$. We then say actors* ***herd*** *on θ.*

In a herding dynamic, the first actors separate or semi-pool until sufficient actions reveal that there is likely a high proportion of θ-types in the group, at which point all actors pool on θ.

The probability of pluralistic ignorance when actors herd will be closely linked to the cutoff $\widehat{\chi}$ and to a related value which I call $\widehat{\chi}^+$. To define $\widehat{\chi}^+$, consider actor 1 of type L's belief over judge 2 at the beginning of the game. If actor 1 of type L knew that the context was equal to $\widehat{\chi}^+$, he would hold the same belief over judge 2 as if he knew that the context was equal to $\widehat{\chi}$ and he observed an extra H signal.[10] Roughly, $\widehat{\chi}^+$ is an extra H-signal above $\widehat{\chi}$. Before giving an intuition of why $\widehat{\chi}^+$ appears, I first state the result.

[10]To find $\widehat{\chi}^+$, set $P(\theta_2 = H \mid \theta_1 = L, \chi = \chi') = P(\theta_2 = H \mid \theta_1 = L, H \text{ signal}, \chi = \widehat{\chi})$ and solve for χ'.

Proposition 2. *Consider a game G with context $\chi \geq 1/2$ and $\pi > (\beta+1)/2\beta$.*

For all contexts below the cutoff $\widehat{\chi}$ there exists a MPBE. All MPBEs consist of a herding dynamic.

The MPBE is unique for groups less or equal to a threshold $I_{unique} \geq 2$. For any $n > 0$ there is a threshold group size such that for all groups greater or equal to that threshold, all actors $i \leq n$ have a unique and pure MPBE strategy.

The MPBE probability of pluralistic ignorance is 0 for groups less or equal to a threshold $I_0 \geq 2$. The MPBE probability of pluralistic ignorance in the subset of realizations of types in which all actors have pure strategy prescriptions is bounded above by $1-\widehat{\chi}$. As the group size grows, the probability of pluralistic ignorance in any MPBE tends to some value that is bounded below by $1-\widehat{\chi} > 0$ and above by $1-\widehat{\chi}^{+}$.

When the context is below the cutoff $\widehat{\chi}$, individuals at the beginning of the game rely heavily on their types to form expectations about the group. Type θ will believe it is relatively more likely that the group was drawn from the population $\psi = \theta$, since that population is mostly likely to have drawn his type. This difference makes actor 1 willing to reveal his type—he believes that many judges will judge him positively for doing so. Actor 1 therefore separates, which gives actor 2 higher incentives to uniquely pool on actor 1's type. Once one action is chosen enough times over the other action, all actors uniquely pool on that action.

Actors learn little about judges' types from the context, but actions themselves reveal information. Pluralistic ignorance arises when the first movers' actions induce individuals to expect that the majority type of judges differs from the true majority type, and therefore herd.

In small groups who follow a herding dynamic, the first movers separate and are a large proportion of the group. Therefore, the first movers' actions provide a lot of information about the composition of the group. For groups small enough that $I \leq I_0$, they provide enough information that there is no pluralistic ignorance.

To show that groups are unique below the threshold I_{unique}, the proof proceeds by showing uniqueness in groups of size 2. Whether uniqueness holds in intermediate-sized groups depends on the parameter values. However, in Section 3.3 I exploit the fact that you can bound the MPBE probabilities of pluralistic ignorance despite the fact that there may be multiple MPBEs.

We can find a lower bound on the information revealed in a herding dynamic conditioning on the realizations of types where actors follow pure strategy prescriptions. This conditioning has a negligible impact as the group size increases: for any number n of first-movers, a large enough group guarantees that those first-movers have a unique and pure MPBE strategies, and the action lead needed to herd is finite. Actors who follow a pure strategy prescription after others have done the same uniquely pool on θ if the type that is not θ expects that there are $(\beta+1)/2\beta$ of the θ-type judges. But then as the group size grows, these actors will uniquely pool on θ if and only if their posterior probability that the population is θ given public information is greater or equal to a value arbitrarily close to the cutoff $\widehat{\chi}$. This is because as the group size grows, the only information that continues to be relevant to determine the composition of the group is information about the population—information about specific individuals' types become negligible. For an arbitrarily large number of actors, the minimum information about the population needed to uniquely pool for an actor who follows a pure strategy is the same as is needed for an actor at beginning of the game.

Notice that this logic also puts an upper bound on the public information when actors follow pure strategy presciptions in a herding dynamic: the posterior probability that the population is θ will not be more than $\widehat{\chi}^{+}$, an extra θ-signal above $\widehat{\chi}$. These bounds are considerably easier to compute than the point probability of pluralistic ignorance in arbitrarily large groups. Further, the bounds are more robust to measurement errors, in the sense that they are the same for any context above the cutoff.

Discussion. An important objective of the model is to make predictions about the probability of pluralistic ignorance and how it varies with the environment, and these predictions would have been sharper with unique MPBEs for all group sizes. Complications in Proposition 2 arise since if an actor does not fully reveal his type, different judges will hold different probabilities that the actor is of type θ. This wedge in beliefs can lead to multiple equilibrium strategy prescriptions:

as an actor's strategy prescription makes it increasingly likely that he is of type θ, the composition of judges that judge him positively may change non-monotonically. A further complication arises from the fact that the strategy prescription an actor follows induces different private information once the actor is a judge. Analyzing equilibrium dynamics therefore requires keeping track of heterogeneous private information of judges that result from different types and different strategy prescriptions as actors. Both of these complications are compounded when actors semi-pool, which may arise even when an actor's strategic prescription is unique. I present a graphical analysis of the difficulties that arise with semi-pooling in Section OD.

I tackle these complications in the proof by focusing on groups that are small enough that semi-pooling remains tractable, and in larger groups by conditioning on realizations of types where all actors follow pure strategy prescriptions. Once again, as the group size grows, semi-pooling arises with a vanishing probability. Further, in Section 3.3 I will provide conservative bonds on the probability of pluralistic ignorance for all groups, even in cases where semi-pooling does arise.

The difficulties I have identified do not arise in standard herding models, where actors are not concerned with how their strategy affects the heterogeneous beliefs other have about them. In those models, actors simply choose an action that they believe best matches the state of the world, and the only role others play is to provide information about that state of the world. A knife-edge scenario, where the actor is indifferent between his preferred action, is not constrained to be played in a way that satisfies judges' beliefs according to some equilibrium logic.

Given the added complications of my model, it is natural to wonder whether we could have obtained similar insights through the simpler setup of a standard herding model. I have already argued in the introduction that pluralistic ignorance is not properly captured by the standard herding model. I now add a further argument: the information revealed in standard herding models is insensitive to group size. In a standard herding model, actors use social information to learn about a state of the world which does not depend on the distribution of true attitudes. Actors will herd on an action once their priors and observed signals make them sufficiently confident in the state of the world. Since the informativeness of a signal does not depend on the group size and the state of the world does not depend on the distribution of true attitudes, the upper bound on what actors learn about the state of the world before herding is not affected by group size, while the lower bound quickly converges as the group size grows. If actors are sufficiently confident in the state of the world (the analogue to a context $\chi > \widehat{\chi}$ in my model), group size has no effect on what actors learn about the state of the world since no information is revealed in equilibrium.

3.3 Maximum and Minimum Probabilities of Pluralistic Ignorance

Here I provide the main result of the paper, as well as several corollaries. The section will address how the probability of pluralistic ignorance varies with the environment, what we can learn about the probability of pluralistic ignorance from actions, and what we can infer about rationality from the probability of pluralistic ignorance.

Despite the fact that the next result is stated in terms of intervals, it will provide unambiguous comparative statics.

Proposition 3. *Consider a game G with a context $\chi \geq 1/2$ and $\pi > (\beta + 1)/2\beta$. There exist thresholds $I_{small} > I_0$ and $I_{large} > I_{small}$ such that:*

In a group $I \leq I_{small}$, if the context is below the cutoff $\widehat{\chi}$, the MPBE probabilities of pluralistic ignorance are lower than for any context outside that range (rectangle with diagonal lines in Figure III).

In a group $I \geq I_{large}$, if the context is sufficiently close to 1 and above the cutoff $\widehat{\chi}$ ($\chi \geq \chi'$ for some $\chi' > \widehat{\chi}$), the MPBE probabilities of pluralistic ignorance are lower than for any context outside that range (rectangle with gridded lines in Figure III).

Consider a group $I \leq I_{small}$. If the context is above the cutoff $\widehat{\chi}$ but sufficiently below χ', the MPBE probabilities of pluralistic ignorance are higher than for any context outside that range (rectangle with horizontal lines in Figure III). The last sentence is true for all group sizes conditioning on the realization of types in which all actors follow pure MPBE strategy prescriptions (rectangle with vertical lines in Figure III).

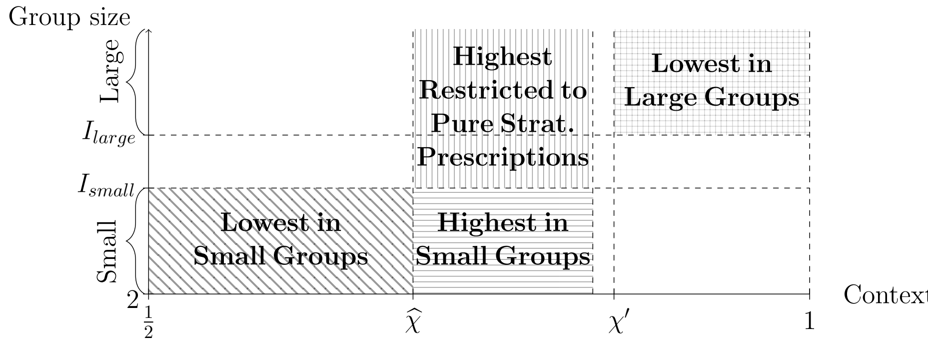

Figure III: Highest and Lowest Probabilities of Pluralistic Ignorance Depend on Group Size as Stated in Proposition 3. The graph indicates the range of contexts where the probability of pluralistic ignorance is highest for small groups (rectangles with horizontal lines), highest conditioning on the realizations of types where actors are following pure strategy prescriptions (rectangles with vertical lines), lowest for small groups (rectangle with diagonal lines) and lowest for large groups (rectangle with gridded lines).

The information provided by each range of contexts explains why different ranges provide the lowest probabilities of pluralistic ignorance in small and large groups. Contexts below the cutoff $\widehat{\chi}$ provide a lot of information about the type of some group members, but provide limited information about the population. The information provided by these contexts is then most revealing about group composition in small groups. Contexts above the cutoff provide a lot of information about the population, but no information about the type of individual members. The information provided by these contexts is then most revealing about the group composition in a large group.

The maximum MPBE probability of pluralistic ignorance in small groups is the probability of drawing a group with majority type L when the context is equal to the cutoff $\widehat{\chi}$. This is also the maximum MPBE probability of pluralistic ignorance for any group conditioning on realizations of types where all actors follow pure strategy prescriptions. To see this, note first that the highest probability of pluralistic ignorance for contexts above the cutoff is given by the context $\widehat{\chi}$ (by Proposition 1). Second, contexts above the cutoff yield higher probabilities of pluralistic ignorance than contexts below the cutoff when the group size is small (Propositions 1 and 2). Third, for realizations of types where all actors follow pure strategy prescriptions, the probability of pluralistic ignorance when the context is equal to the cutoff tends to the upper bound of the probability of pluralistic ignorance in contexts below the cutoff (Propositions 1 and 2).[11] Recall that by Proposition 2, the proportion of realization of types where all actors follow pure strategy prescriptions goes to 1 as the group size increases.

Note the cutoff $\widehat{\chi}$ approaches $1/2$ from above as the weight on expected judgment β becomes arbitrarily large and for a precision π equal to the cutoff—by L'Hôpital's rule applied to equation (2). As mentioned in the introduction, this implies that the probability of pluralistic ignorance approaches $1/2$ from below for arbitrarily large values of β and of the group size I, and for values of π arbitrarily close to $1/2$. Although the numerical exercise is meant to show how large the probability of pluralistic ignorance can be in the model, I conjecture that it is empirically rare to observe situations in which actors put full weight on conforming to expected judgments.

I now establish comparative statics that apply to realizations of types where all actors follow unique MPBE strategy prescriptions. This restriction does not affect group sizes less or equal to I_{unique}, and is negligible in large groups (by Proposition 2):

Corollary 1. *Consider a game G with context $\chi \geq 1/2$ and $\pi > (\beta+1)/2\beta$, and consider the subset of realization of types $\widehat{\Theta}'$ where actors follow unique MPBE strategy prescriptions.*

The MPBE probability of pluralistic ignorance conditioning on $\widehat{\Theta}'$ weakly increases as the weight on expected judgment β increases, and the expected proportion of actors who choose H in the MPBE weakly increases in the context χ.

The result regarding the weight on expected judgment β follows from two facts. First, that an increase in β decreases the cutoff $\widehat{\chi}$. Second, in a herding dynamic actors need less information to pool with a higher β, and reveal less information if they semi-pool. Therefore, for all contexts, an increase in β decreases the information that is revealed about the group. The result regarding the context χ follows from the fact that with a larger χ, actors need more information revealed by others' types in order to choose L, and less information to choose H.

Non-monotonicity in I, χ, π. When actions follow a herding dynamic, the probability of

[11] The presence of semi-pooling strategy prescriptions could in principle yield probabilities of pluralistic ignorance higher than $1-\widehat{\chi}$, since actors may reveal more information before uniquely pooling than with pure strategy prescriptions as exemplified in Section OB.

pluralistic ignorance does not change monotonically or continuously with the group size I, the precision π or the context χ (even conditioning on realizations of types where all actors follow pure strategy prescriptions). Examples are given in Section OB. The lack of monotonicity and the discontinuity is due to the discreteness of the group sizes and of the information provided by revealing a signal. Essentially, although a change in the context makes it easier to herd on H and harder to herd on L, the change may affect the integer of the lead of action H needed to herd on H without affecting the integer of the lead of action L needed to herd on L, or vice versa. A change in the group size I may also change the contexts where the thresholds for herding change. But changes in one threshold for herding but not the other may affect the probability of pluralistic ignorance in different directions. An increase in the precision, on the other hand, may decrease the cutoff $\widehat{\chi}$ from above the context to below the context, leading to a discontinuous change in the probability of pluralistic ignorance.

Sizing small and large groups. As argued above, computing the probability of pluralistic ignorance with contexts below the cutoff $\widehat{\chi}$ is complicated by semi-pooling strategy prescriptions and multiple MPBEs. Nevertheless, we can still bound the probability of pluralistic ignorance since we know that actors follow a herding dynamic. Here I derive easy-to-compute upper and lower bounds on the probability of pluralistic ignorance, with details and a program for computing these bounds developed in Section OC. The upper bound $\overline{P}$ will allow me to bound I_{large}. To define $\overline{P}$, let $\widehat{\theta}$ be the strict majority type in the group. Then, $\overline{P}(I) \equiv P(\widehat{\theta} = L, \theta_1 = H) + P(\widehat{\theta} = H, \theta_1 = L)$. The term $\overline{P}$ captures the probability of pluralistic ignorance induced by a strategy profile in which actor 1 separates, all other actors pool on actor 1's type. This is an upper bound on the probability of pluralistic ignorance since the revelation of information about the groups' types in a herding dynamic is minimized.

Similarly, we can derive a lower bound $\underline{P}$ on the probability of pluralistic ignorance, with which I will bound I_{small}. This lower bound is obtained from assuming that pluralistic ignorance only arises when there is the largest possible run of type θ signals and the majority type is of type $-\theta$. The largest possible run $\mathcal{N}$ is obtained from Lemma 5, which gives the maximum threshold

$\overline{N}(a, h_i)$ for a given set of parameters β, π and χ. This specification is a conservative lower bound, but focuses on runs where semi-pooling would not arise, and in which strictly more information is revealed than in any other herding dynamic. Then, $\underline{P}(I) \equiv P(\text{run of } \mathcal{N}\ H-\text{signals}, \widehat{\theta} = L) + P(\text{run of } \mathcal{N}\ L-\text{signals}, \widehat{\theta} = H)$. By symmetry of the game, neither $\overline{P}$ nor $\underline{P}$ depend on χ.

The following Corollary uses the fact context $\chi = 1$ yields a lower bound for the probability of pluralistic ignorance when the context is above the cutoff $\widehat{\chi}$.

Corollary 2. *Consider a game G with $\chi \in [1/2, \widehat{\chi})$ and $\pi > (\beta + 1)/2\beta$, and let $P(I)$ be the probability of pluralistic ignorance when the context χ is one and group size is I.*

Let $\tilde{I}$ be the smallest group size such that $P(\tilde{I}) > \overline{P}(\tilde{I})$. Then $\tilde{I} \leq I_{small}$.

Let $\widehat{I}$ be the smallest group size such that $P(\widehat{I}) < \underline{P}(\widehat{I})$, and consider an odd-numbered group size $I \geq \widehat{I}$. Conditional on realizations of types in which actors follow pure MPBE strategy prescriptions, $\widehat{I} \geq I_{large}$.

The last paragraph of Corollary 2 focuses on odd-numbered groups to ensure monotonicity of $P(I)$ and $\underline{P}$.[12] This is not very restrictive since with large enough groups there is a negligible change in the probability of pluralistic ignorance as the group size increases. Similarly, conditioning on realizations of types where all actors follow a pure strategy prescription has a negligible impact for large groups (once again using Proposition 2). Therefore, the Corollary gives a good approximation to bounding large groups.

We can then use Corollary 2 to provide a lower bound on the largest small group and an upper bound on the smallest large group, for given parameter values. For example, if $\beta = 6$, $\pi = .6$ and $\chi = 0.7$, then small groups include groups of size 9, and large groups include groups of size 509. If $\beta = 3$, $\pi = .7$ and $\chi = 0.6$, small groups include groups of size 4, and large groups include groups of size 87.

Finally, note that obtaining the exact values of I_{large} and I_{small} would require a comparison

[12]The probability of pluralistic ignorance when all pool on H does not decline monotonically in the group size I. This is partly due to the differences in the probabilities between groups of even and odd-numbered sizes, although there may also be non-monotonicity among the set of even-numbered groups. Nevertheless, in Fernández-Duque (2021b) I show that the probability declines monotonically in group size among odd-numbered groups.

akin to that in Corollary 2, replacing the bounds with the precise probabilities, and keeping in mind that the probabilities for contexts above and below the cutoff $\widehat{\chi}$ may cross more than once due to the non-monotonicity in group size I.

Probabilities of actions and of pluralistic ignorance. What is the relationship between the probability of pluralistic ignorance and the probability actors choose one of the actions? The relationship is monotonic in strategic complementarities models of pluralistic ignorance, where pluralistic ignorance is equated to coordinating on the inefficient action. However, the relationship is not monotonic in my model.

Corollary 3. *Consider a game G with context $\chi > 1/2$ and $\pi > (\beta + 1)/2\beta$.*

For a group $I \leq I_{small}$ or $I \geq I_{large}$, there are changes in the context that increase the MPBE probability of pluralistic ignorance and increase the expected proportion of actors that choose H, and there are changes in the context that increase the MPBE probability of pluralistic ignorance and decrease the expected proportion of actors that choose H.

By Corollary 1, we know that the expected proportion of actors who choose H increases with the context. In particular, the expected proportion is at its maximum when the context is close enough to one (that is, above the cutoff $\widehat{\chi}$), where all actors choose H. The expected proportion is at its minimum when the context is close enough to zero (below $1 - \widehat{\chi}$), where no actor chooses H.

However, there is not a monotonic relationship between the probability of pluralistic ignorance and the context. Suppose the group was large or small and the context started above the cutoff $\widehat{\chi}$, at the range of values where the probability of pluralistic ignorance is higher than for any value outside that range (by Proposition 3). Now consider a shock to the context such that it ended up below $1-\widehat{\chi}$, at the range of values where the probability of pluralistic ignorance is lower than where it started (this range exists by Proposition 3 again, exploiting the symmetry of the model). This shows that a decrease in the probability of pluralistic ignorance can be accompanied by a decrease in the context. But by symmetry of the model, there also exists a decrease in the probability of pluralistic ignorance associated to an increase in the context—for example with a change from a context in intermediate values of the range $[0, 1/2]$, where all actors pool on L and the probability of pluralistic ignorance is in the range that takes on the highest values, to a context sufficiently close to one, in a range where the probability of pluralistic ignorance is not the highest.

How rational is pluralistic ignorance? The underlying mechanism that causes pluralistic ignorance has been attributed to behavioral and rational mechanisms by different authors across disciplines (compare rational explanations in Kuran, 1997, Centola et al., 2005, or Chwe, 2013 with non-rational explanations reviewed in Shamir and Shamir, 2000 and in Kitts, 2003). Some authors consider irrationality an essential ingredient of pluralistic ignorance (e.g. Bicchieri, 2005). However, my model has shown that pluralistic ignorance can be an equilibrium outcome without such underestimation. By observing whether the probability of pluralistic ignorance follows the predictions of the model, for instance whether it falls within the predicted bounds for a given group size and a given context, we can empirically test whether a rational model is sufficient to explain the pluralistic ignorance we observe.

Comments on the assumption that $\pi > (\beta+1)/2\beta$. The assumption guarantees that, holding the other parameters constant, the proportion of θ-types in population θ is sufficiently high that an actor would pool on θ if he had strong enough beliefs that the population is θ.

Since I am interested in studying the probability of pluralistic ignorance, the assumption is valuable for two reasons. First, since two ranges of contexts exist (below and above the cutoff $\widehat{\chi}$), the assumption allows me to contrast the probability of pluralistic ignorance across both ranges. Second, it is conducive to the existence of pluralistic ignorance in equilibrium. If actors would not pool on θ even if it were common knowledge that the population was θ, then in arbitrarily large groups the probability of pluralistic ignorance is arbitrarily close to 0. Indeed, the proportion of θ-types in population θ would equal π with an arbitrarily high probability for arbitrarily large histories (by the law of large numbers), which would not be sufficient for actors to pool on θ.

4 Existing Evidence

Here I consider the existing evidence relevant for assessing the predictions of the model, and the larger body of evidence testing the behavioral assumptions of the model. I end with directions for future research.

To the best of my knowledge, there are few studies about the probability of pluralistic ignorance.[13] However, pluralistic ignorance is a phenomenon with potentially important social consequences. As others have argued, it can explain the persistence of inefficient norms and abrupt social change . In order to understand how much of social behavior can be explained by pluralistic ignorance, we must start by understanding its prevalence.

Laboratory evidence. Although I have not found laboratory evidence studying the probability of pluralistic ignorance in the way I have defined it, the closest relevant findings I know of come from recent work by Smerdon et al. (2020), Andreoni, Nikiforakis, and Siegenthaler (2019), Andreoni et al. (2020) and Duffy and Lafky (2020). They all study settings in which actors play coordination games with preference changes. Smerdon et al. (2020), Andreoni et al. (2019) and Duffy and Lafky (2020) show that actors may become stuck coordinating on an inefficient outcome after preferences have changed, while Andreoni et al. (2020) accurately predict when actors move away from this inefficient outcome. Smerdon et al. (2020) show that the inefficient outcome arises only when there is incomplete information about others' preferences, while Andreoni et al. (2019) and Duffy and Lafky (2020) show that inefficient outcomes are more likely when the utility from coordinating is higher.

Note that although some of these authors refer to the inefficient outcome as pluralistic ignorance (Smerdon et al., 2020, Duffy and Lafky, 2020), the reason for being stuck in the inefficient outcome differs conceptually from pluralistic ignorance in my model: actors know that others' preferences are changing, but still are better off choosing the inefficient outcome than not coordi-

[13]The empirical literature has mostly focused on documenting specific instances of pluralistic ignorance. This opens the possibility that pluralistic ignorance is not very prevalent, or that it is difficult to identify across a range of issues—as if it is only worth studying pluralistic ignorance when the researcher has a special insight into a specific instance of its occurrence. This interpretation echoes an argument made by Kuran (1989) in his seminal paper, that rapid social change is hard to predict in part because of the difficulty of identifying pluralistic ignorance.

nating. Moreover, in these models conformity is the weight in their payoff on coordinating, not the weight on being judged by others. Despite these differences, their findings that incomplete information leads to the inefficient outcome is broadly consistent with my model, and their findings that the weight on social preferences makes the inefficient outcome more likely echo Corollary 1.

Other laboratory evidence has tested for key features of my model more closely. For example, Bursztyn et al. (2017) take advantage of Donald Trump's 2016 election to the U.S. presidency to test how actors react to judgments they expect others to make, and how judges judge others. They show that when individuals think they will be judged by observers, they act according to how they think they will be judged best: in public, their answers change if they believe most of the observers support Trump or not. Further, a separate experiment shows that observers' judgments take into account whether decision-makers are trying to manage the observers' impressions, which affects whether they materially benefit or punish the decision maker. Relatedly, there is a large literature on conformity from psychology (Asch, 1956 is a classic reference, and for a recent review see Hodges, 2017). These studies typically put subjects in a situation where there is pressure to conform to expected judgments, and have provided robust evidence that a subject does so. For example, Willer, Kuwabara, and Macy (2009) show that when subjects make private judgments over other subjects' ranking of wines or obscure texts, they have a more positive judgment of those whose rankings are more similar to their preferences. This is true whether or not the subjects making the judgments had previously conformed in their own ranking.[14]

Field evidence. One exception to the dearth of research focused on the probability of pluralistic ignorance is Shamir and Shamir (2000), who studied public opinion in Israel over 24 public opinion topics. Their design tests how pluralistic ignorance in large groups varies with information about others' opinions. They find that topics that were more "publicly visible" were less likely to have pluralistic ignorance. Their measures of public visibility arguably capture variation in the context, and therefore provide some support for the part of Proposition 3 regarding large groups. To give

[14]To measure whether subjects previously conformed, they put them in a situation in which confederates express an attitude the subject clearly does not share: out of three otherwise identical wines, confederates choose the one that has added vinegar, or claim to understand an unintelligible text. Then they measure whether subjects expresses the same attitude as the confederates.

some examples of the way they capture public visibility of a topic, they use media experts' opinion on how much information the public had access to, or whether individuals could infer obscure attitudes from better known differences in opinions—such as inferring the distribution of several opinions regarding the Israel-Arab conflict from the better known left-right political divide.

Future work. Shamir and Shamir (2000) provide a natural starting point for testing the large-group results of the paper. More work needs to be done, both to increase the units of observation and to better address the methodological concerns in studying pluralistic ignorance (a good example of recent developments in methodology for studying pluralistic ignorance is Bursztyn et al., 2020).

A laboratory environment provides a natural setting to study the small-group results of Proposition 3. For example, a researcher could elicit subjects' preferences at baseline, allow subjects to exchange information about their preferences with others in a group, elicit beliefs over others' preferences and measure the percentage of groups with pluralistic ignorance. Within this framework, the experimenter could vary the context by manipulating the *ex ante* information individuals have about others' preferences.

5 Conclusion

In this paper I introduced a model of dynamic attitude expression with which I derived the equilibrium probability of pluralistic ignorance. The model provides several theoretical contributions.

First, I provide a formal definition which captures what I have argued are key features of pluralistic ignorance: (a) most individuals do not express their true attitude, (b) most believe most others do express their true attitude, and (c) it is the overestimation of how many express their true attitude that motivates an individual to not express his true attitude. I then showed how different dynamics of social learning may lead to pluralistic ignorance.

Second, the model yields upper and lower bounds on the probability of pluralistic ignorance across group sizes, providing theoretical guidance to the question of how likely pluralistic ignorance is to arise, and a way to distinguish between rational and behavioral theories. The model provides a theory of how pluralistic ignorance depends on group size: the probability of pluralistic ignorance is lowest in small groups when there is high uncertainty about the population from which true attitudes are drawn, and lowest in large groups when there is low uncertainty. The probability of pluralistic ignorance is highest in small and large groups when there is intermediate second-order uncertainty. A Corollary helps bound the size of small and large groups.

Third, the theory sheds light on the parameters that affect the probability of pluralistic ignorance. The paper identifies monotonic and non-monotonic relationships, which helps to understand the scope of testable predictions. The uncertainty about the population from which true attitudes are drawn has a non-monotonic impact on the probability of pluralistic ignorance, but by the main result, meaningful comparisons can still be made across ranges of this uncertainty. Furthermore, since this uncertainty about the population has a monotonic impact on the probability actors choose an action, it follows that there is no monotonic impact between the probability of pluralistic ignorance and the probability of choosing a certain action. Group size generally has a non-monotonic impact, except among odd-numbered groups with low uncertainty. The weight on expected judgments, on the other hand, does have a monotonic impact.

Past work has established that pluralistic ignorance is at the heart of many important social phenomena: the persistence in conservative social norms such as segregation, religious observance or political correctness, regime-changing revolutions such as the Arab Spring or the fall of Communist regimes, or surprising elections such as Trump's 2016 election or Brexit, and the adoption of behavior and beliefs that affects others such as climate change beliefs, alcohol consumption, tax compliance, or support for female labor force participation. This important work has established that instances of pluralistic ignorance do happen, and that its occurrence is consequential. My work takes a step towards quantifying how often it happens and when it is more likely to occur.

References

S Nageeb Ali and Roland Bénabou. Image versus information: Changing societal norms and optimal privacy. *American Economic Journal: Microeconomics*, 12(3):116–64, 2020.

S Nageeb Ali and Navin Kartik. Herding with collective preferences. *Economic Theory*, 51(3): 601–626, 2012.

James Andreoni, Nikos Nikiforakis, and Simon Siegenthaler. Social change and the conformity trap. Technical report, Technical report, Mimeo, 2019.

James Andreoni, Nikos Nikiforakis, and Simon Siegenthaler. Predicting social tipping and norm change in controlled experiments. Technical Report No. w27310, National Bureau of Economic Research, 2020.

Solomon E Asch. Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological monographs: General and applied*, 70(9):1, 1956.

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

Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678, 2006.

Roland Benabou and Jean Tirole. Laws and norms. *National Bureau of Economic Research*, (No. w17579), 2012.

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

Cristina Bicchieri. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press, 2005.

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

Leonardo Bursztyn and Robert Jensen. Social image and economic behavior in the field: Identifying, understanding, and shaping social pressure. *Annual Review of Economics*, 9:131–153, 2017.

Leonardo Bursztyn, Georgy Egorov, and Stefano Fiorin. From extreme to mainstream: How social norms unravel. *National Bureau of Economic Research*, (No. w23415), 2017.

Leonardo Bursztyn, Alessandra L González, and David Yanagizawa-Drott. Misperceived social norms: Women working outside the home in saudi arabia. *American Economic Review*, 2020.

Damon Centola, Robb Willer, and Michael Macy. The emperor's dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, 110(4):1009–1040, 2005.

Yeon-Koo Che, Wouter Dessein, and Navin Kartik. Pandering to persuade. *American Economic Review*, 103(1):47–79, 2013.

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

Michael Suk-Young Chwe. Structure and strategy in collective action. *American journal of sociology*, 105(1):128–156, 1999.

Michael Suk-Young Chwe. *Rational ritual: Culture, coordination, and common knowledge*. Princeton University Press, 2013.

Robyn M Dawes. Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, 25(1):1–17, 1989.

Stefano DellaVigna, John A List, and Ulrike Malmendier. Testing for altruism and social pressure in charitable giving. *The quarterly journal of economics*, 127(1):1–56, 2012.

John Duffy and Jonathan Lafky. Social conformity under evolving private preferences. *unpublished manuscript*, 2020.

Tore Ellingsen, Magnus Johannesson, et al. Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3):990–1008, 2008.

Dirk Engelmann and Martin Strobel. The false consensus effect disappears if representative information and monetary incentives are given. *Experimental Economics*, 3(3):241–260, 2000.

Dirk Engelmann and Martin Strobel. Deconstruction and reconstruction of an anomaly. *Games and Economic Behavior*, 76(2):678–689, 2012.

Mauricio Fernández-Duque. Pluralistic ignorance and collective action. *Unpublished manuscript, available at scholar.harvard.edu/duque*, 2021a.

Mauricio Fernández-Duque. The probability of pluralistic ignorance. *Available at SSRN:http://ssrn.com/abstract=3992145*, 2021b.

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

Matthew Gentzkow and Jesse M Shapiro. Media bias and reputation. *Journal of political Economy*, 114(2):280–316, 2006.

Ben Golub and Evan Sadler. Learning in social networks. In *The Oxford Handbook of the Economics of Networks*. 2016.

Bert H Hodges. Conformity and divergence in interactions, groups, and culture. *The Oxford Handbook of Social Influence*, page 87, 2017.

Daniel Katz and Floyd H Allport. Students' attitudes. *Syracuse, NY: Craftsman*, page 152, 1931.

James A Kitts. Egocentric bias or information management? selective disclosure and the social roots of norm misperception. *Social Psychology Quarterly*, pages 222–237, 2003.

Timur Kuran. Sparks and prairie fires: A theory of unanticipated political revolution. *Public choice*, 61(1):41–74, 1989.

Timur Kuran. *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press, 1997.

Gilat Levy. Decision making in committees: Transparency, reputation, and voting rules. *American economic review*, 97(1):150–168, 2007.

Matto Mildenberger and Dustin Tingley. Beliefs about climate beliefs: The importance of second-order opinions for climate politics. *British Journal of Political Science*, pages 1–29, 2017.

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

Hubert J O'Gorman. Pluralistic ignorance and white estimates of white support for racial segregation. *Public Opinion Quarterly*, 39(3):313–330, 1975.

Canice Prendergast. A theory of "yes men". *The American Economic Review*, pages 757–770, 1993.

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

Lee Ross, David Greene, and Pamela House. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, 13(3): 279–301, 1977.

Jacob Shamir and Michal Shamir. *The anatomy of public opinion*. University of Michigan Press, 2000.

Dirk Sliwka. Trust as a signal of a social norm and the hidden costs of incentive schemes. *American Economic Review*, 97(3):999–1012, 2007.

David Smerdon, Theo Offerman, and Uri Gneezy. 'everybody's doing it': on the persistence of bad social norms. *Experimental Economics*, pages 1–29, 2020.

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

Cass R Sunstein. *Conformity: the power of social influences*. NYU Press, 2019.

Bauke Visser and Otto H Swank. On committees of experts. *The Quarterly Journal of Economics*, 122(1):337–372, 2007.

Michael Wenzel. Misperceptions of social norms about tax compliance: From theory to intervention. *Journal of Economic Psychology*, 26(6):862–883, 2005.

Robb Willer, Ko Kuwabara, and Michael W Macy. The false enforcement of unpopular norms. *American Journal of Sociology*, 115(2):451–490, 2009.

Appendices

Appendices

A Lemmas

There are some preliminary results and concepts that are needed for the rest of the results. I begin by characterizing actors' and judges' beliefs in terms of primitives and the information revealed through the history of play (Lemma 1). I then describe the relationship between the actor's and judges' beliefs (Lemma 2). Lemma 3 characterizes the contexts close to $1/2$ that lead to herding dynamics and the contexts close to 1 that lead all to pool on H. The next three lemmas are key for showing that pure strategies are followed in large groups. Lemma 4 shows how, as the group size grows, actors' expected proportion of judges and judges' beliefs over actors are increasingly similar. Lemma 5 shows that the threshold for herding in a herding dynamic is finite in large groups. Lemma 6 shows that if past actors have played pure strategy profiles, the distance between actors' and judges' equilibrium beliefs can be made arbitrarily small as the group size increases.

For brevity, from this point on I denote $-x$ as the element other than $x \in \{H, L\}$ in the set.

I call k a *withholder* at history h_i if one of two conditions hold: (a) k has already made a decision ($k < i$) and he pooled, or (b) k has not made a decision ($k > i$). I call k a *revealer* at history h_i if k has already made a decision ($k < i$) and separated.

To calculate an actor's expected judgment, we need to know how he forms his beliefs over how he is judged by others. A judge j judges others whether or not she has already acted. Therefore, actor i's beliefs about how j will judge him will depend on what j has disclosed about her type through her actions, captured by j's *disclosure* at history h_i (in equilibrium σ), $\theta_j^o : \mathcal{H}_i \to [-1, 1]$. If $j < i$ chose a at history h_i, her disclosure is defined by:

$$\left(\frac{\pi}{1-\pi}\right)^{\theta_j^o} = \frac{P(a_j = a \mid \theta_j = H, h_i)\pi + P(a_j = a \mid \theta_j = L, h_i)(1-\pi)}{P(a_j = a \mid \theta_j = H, h_i)(1-\pi) + P(a_j = a \mid \theta_j = L, h_i)\pi} = \frac{P(a_j \mid h_i, \psi = H)}{P(a_j \mid h_i, \psi = L)}$$

If $j > i$, her disclosure is 0. The disclosure takes the value -1 if j has revealed type L, the value +1 if j has revealed type H, and the value 0 if j is a withholder. The disclosure takes a value in $(-1, 0)$ or in $(0, 1)$ for intermediate cases. For example, if it was more likely for $j < i$ of type L to choose a_j than j of type H, then $\theta_j^o(h_i) \in (-1, 0)$. Note the disclosure strictly increases in the ratio of the probabilities that types H and L chose a.

Define an individual's *type duple* as the pair (θ_j^o, θ_j). If j is a withholder of type θ, I will simplify notation by writing her type duple as $w\theta$. If j is a revealer, I will write $r\theta$.

When it is useful, I will explicitly condition j's belief on her disclosure θ_j^o—for example, I will write $P(\theta_i = H \mid h_i, (\theta_j^o, \theta_j))$. Although the conditioning on θ_j^o is implicit in the expression $P(\theta_i = H \mid h_i, \theta_j)$, it will sometimes make it easier to keep track of the information j has when forming beliefs.

The judges' and actors' decisions will depend on their beliefs. When actor i pools, judge j's judgment of i will depend on her beliefs over i's most likely type. In turn, actor i's decision will depend on his belief over the distribution of judges' types. Although what an actor directly cares about is the beliefs judges form about his type given their type, it will turn out that the belief over the distribution of judges' types is a key input. To capture what an individual k knows about the group, let the *disclosure lead* of action H be $\Delta^o(h_i, \theta_k) \equiv \sum_{l \neq k} \theta_l^o(h_i) + (2 \times \mathbb{1}\{\theta_k = H\} - 1)$, or the difference between the type H and type L signals observed by individual k of type θ_k at history h_i for some equilibrium σ. The disclosure lead of action L is $-\Delta^o$.

Lemma 1. *Judge j will believe a withholder i is of type H with probability at least z, or $P(\theta_i = H \mid h_i, a_i, \theta_j^o, \theta_j) > z$, if and only if:*

$$\tilde{\pi}^{\Delta^o} \left[P(\theta_i = H \mid h_i, a_i, \psi = H) - z\right] - \underset{\sim}{\chi} \left[z - P(\theta_i = H \mid h_i, a_i, \psi = L)\right] > 0 \Leftrightarrow$$

$$r(\Delta^o) P(\theta_i = H \mid h_i, a_i, \psi = H) + (1 - r(\Delta^o)) P(\theta_i = H \mid h_i, a_i, \psi = L) > z \tag{3}$$

where $\tilde{\pi} = \pi/(1-\pi)$, $\underset{\sim}{\chi} = (1-\chi)/\chi$, and $r(\Delta^o) = \tilde{\pi}^{\Delta^o}/(\tilde{\pi}^{\Delta^o} + \underset{\sim}{\chi})$.

Actor i will believe the average judge is of type H with probability at least z, or $\mathbb{E}_{j \neq i}\big(P(\theta_j = H \mid h_i, \theta_i^o, \theta_i)\big) > z$, if and only if:

$$\tilde{\pi}^{\Delta^o} \left[\mathbb{E}_{j \neq i}\big(P(\theta_j = H \mid h_i, \psi = H)\big) - z\right] - \underset{\sim}{\chi} \left[z - \mathbb{E}_{j \neq i}\big(P(\theta_j = H \mid h_i, \psi = L)\big)\right] > 0 \Leftrightarrow$$

$$r(\Delta^o)\mathbb{E}_{j\neq i}(P(\theta_j = H \mid h_i, \psi = H)) + (1 - r(\Delta^o))\mathbb{E}_{j\neq i}(P(\theta_j = H \mid h_i, \psi = L)) > z \qquad (4)$$

Proof. Write j's posterior over i's type at period i using the law of iterated expectations:

$$P(\theta_i = H \mid h_i, a_i, \theta_j) = \sum_{\hat{\psi}\in\{H,L\}} P(\psi = \hat{\psi} \mid h_i, a_i, \theta_j)P(\theta_i = H \mid \psi = \hat{\psi}, h_i, a_i)$$

Note that $P(\theta_i = H \mid \psi, h_i, a_i, \theta_j) = P(\theta_i = H \mid \psi, h_i, a_i)$, since once we condition on the population ψ, the only reason to condition on history (h_i, a_i) is to determine i's optimal choice given his strategy, which is publicly known in equilibrium. We can apply Bayes' rule to obtain

$$\frac{\sum_{\hat{\psi}\in\{H,L\}} P(\psi = \hat{\psi}, h_i, a_i, \theta_j)P(\theta_i = H \mid \psi = \hat{\psi}, h_i, a_i)}{P(h_i, a_i, \theta_j)}$$

By Bayes' rule again, $P(\psi, h_i, a_i, \theta_j) = P(\psi)P(h_i, a_i, \theta_j \mid \psi)$. By the law of iterated expectations, $P(h_i, a_i, \theta_j) = \sum_{\hat{\psi}\in\{H,L\}} P(\psi = \hat{\psi})P(h_i, a_i, \theta_j \mid \psi = \hat{\psi})$. Since draws are independent, $P(h_i, a_i, \theta_j \mid \psi) = P(h_i, a_i \mid \psi)P(\theta_j \mid \psi)$. This equality again uses the fact that, conditional on ψ, private information is irrelevant for calculating the probability of a given history (h_i, a_i). Further, the action of actor 1 does not depend on others' actions, although others' actions depend on the first actors' actions: $P(h_i, a_i \mid \psi) = P(a_1 \mid \psi)P(a_2, ..., a_i \mid a_1, \psi)$. But conditional on ψ and a_1, actor 2's action does not depend on others' actions. Iterating this argument, we get

$$\frac{\sum_{\psi\in\{H,L\}} P(\psi)P(\theta_j \mid \psi)P(\theta_i = H \mid \psi, h_i, a_i)\prod_{k\in\{1,...,i-1\}} P(a_k \mid h_k, \psi)}{\sum_{\psi\in\{H,L\}} P(\psi)P(\theta_j \mid \psi)\prod_{k\in\{1,...,i-1\}} P(a_k \mid h_k, \psi)}$$

Setting $P(\theta_i = H \mid h_i, a_i, \theta_j) > z$, noting that $P(\theta_i = H \mid \psi, h_i, a_i) = P(\theta_i = H \mid \psi)$ since i is a withholder at period i, and rearranging terms, we get

$$\frac{\prod_{k\in\{1,...,i-1\}} P(a_k \mid h_k, \psi = H)P(\theta_j \mid \psi = H)}{\prod_{k\in\{1,...,i-1\}} P(a_k \mid h_k, \psi = L)P(\theta_j \mid \psi = L)}(\pi - z) > \frac{1-\chi}{\chi}(z - (1 - \pi))$$

We can then apply the definition of disclosures to get the first part of the Lemma.

We can follow an analogous set of steps to obtain that $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \theta_i)\big)$ is equal to

$$\frac{\sum_{\psi\in\{H,L\}} P(\psi)P(\theta_i \mid \psi)\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \psi)\big) \prod_{k\in\{1,\dots,i-1\}} P(a_k \mid h_k, \psi)}{\sum_{\psi\in\{H,L\}} P(\psi)P(\theta_i \mid \psi) \prod_{k\in\{1,\dots,i-1\}} P(a_k \mid h_k, \psi)}$$

By setting $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \theta_i)\big) > z$, noting that $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \theta_i, \psi)\big)$ is equal to $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \psi)\big)$ since private information is not necessary for inferring the probability of others' type given the population ψ, rearranging the terms, and applying the definition of disclosures, we get the second part of the Lemma. □

The expressions in Lemma 1 characterize the beliefs that shape actors' and judges' choices, in terms of primitives and the disclosure lead. The term $\tilde{\pi}^{\Delta^o}$ is the likelihood ratio of the precision π, to the power of the disclosure lead Δ^o. This term captures the information an individual has about the population based on the signals she or he has observed. The expressions depend positively on the context χ, which appears via its inverse likelihood ratio $\underset{\sim}{\chi}$. The judges' and actors' beliefs described in the Lemma only differ according to what they are forming beliefs over. Actors are concerned about the expected type of all judges, whereas judges of actor i are concerned about the expected type of i.

Although a judge's own disclosure is not an argument in (3) and an actor's own disclosure is not an argument in (4), their disclosure will affect how strong their beliefs are relative to others in a group.

The next Lemma shows how beliefs vary with the disclosure lead and the context.

Lemma 2. *The terms $P(\theta_i = H \mid h_i, a_i, \theta_j^o, \theta_j)$ and $\mathbb{E}_{j\neq i}P(\theta_j = H \mid h_i, \theta_i^o, \theta_i)$ increase in χ.*

Increase $\sum_{l\neq k} \theta_l^o$ by increasing the disclosure of some individual l. Then $P(\theta_i = H \mid h_i, a_i, \theta_j^o, \theta_j)$ and $\mathbb{E}_{j\neq i}P(\theta_j = H \mid h_i, \theta_i^o, \theta_i)$ increase.

The probability that judges give to an actor being of type H is increasing in the disclosure for judges of type L ($\partial P(\theta_i = H \mid h_i, a_i, \theta_j^o, \theta_j = L)/\partial\theta_j^o > 0$), decreasing in the disclosure for judges of type H ($\partial P(\theta_i = H \mid h_i, a_i, \theta_j^o, \theta_j = H)/\partial\theta_j^o < 0$), and equal for revealers of type H and of type L.

Suppose $\chi \geq 1/2$. If the disclosure lead is equal to 0, revealer judges believe that a withholder actor is most likely of type H: $P(\theta_i = H \mid \Delta^o = 0, (\theta_j^o, \theta_j) = r\theta) \geq 1/2$.

Proof. The Lemma follows from Lemma 1.

An increase in χ increases $P(\theta_i = H \mid h_i, a_i, \theta_j^o, \theta_j)$ and $\mathbb{E}_{j\neq i} P(\theta_j = H \mid h_i, \theta_i^o, \theta_i)$ straightforwardly.

An increase in $\sum_{l\neq j} \theta_l^o(h_i)$ increases $r(\Delta^o)$ in (3) and (4). The sum of disclosures $\sum_{l\neq i} \theta_l^o(h_i)$ also affects g_i through $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \psi)\big)$, which we can rewrite as follows:

$$\sum_{x\in[-1,1]} P(\theta_{j\neq i}^o = x \mid h_i) P(\theta_{j\neq i} = H \mid \theta_j^o = c, \psi)$$

The term $P(\theta_{j\neq i}^o = x \mid h_i)$ is the probability that judge j of actor i is of disclosure x, and does not depend on ψ because PBE strategies cannot condition on an unobserved parameter. The term $P(\theta_{j\neq i} = H \mid \theta_j^o = x, \psi)$ is the probability that judge j of actor i is of type H when the population is known to be ψ and the judge is of disclosure x. Once the population is known, history does not provide information about a withholder's type. Notice that the sum runs through an uncountable set, which is valid since there are only a countable number of disclosures for a given group.

The expectation $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \psi)\big)$ is a weighted sum of revealers, withholders, and all disclosures with in-between values. Revealers of type H receive a weight of 1, withholders receive a weight between 0 and 1, and revealers of type L receive a weight of 0. In-between disclosures have an in-between weight. For example, if the action taken by judge $j < i$ on the history of play was more likely chosen by type H than by type L, the weight is greater than that of withholders but less than 1. Therefore, the expectation would increase if and only if $\sum_{l\neq k} \theta_l^o(h_i)$ increases—by replacing an individual with a lower disclosure with one with a higher disclosure (e.g. by replacing a revealer of type L with a withholder). This also increases $\Delta^o(h_i, \theta_l)$. Since (4) increases in $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \psi)\big)$ for $\psi \in \{H, L\}$, the result follows.

Now I will show how types and disclosures affect beliefs that the actor is of type H. The only difference between what judges know is the private information each has about their own type, which depends on their disclosure. The higher the disclosure of a judge j of type L, the more others believe j is of type H. Then the stronger the belief j has that the population and his own type are L relative to another individual's belief. Similarly, the lower the disclosure of a judge of type H, the stronger the belief the judge has that the population and his own type are H relative to another individual's belief. The result follows.

I now show that $P(\theta_i = H \mid \Delta^o = 0, (\theta_j^o, \theta_j) = r\theta) \geq 1/2$. Note

$$P(\theta_i = H \mid h_i, a_i, \theta_j^o, \theta_j) - 1/2 = (\pi - .5)\left[\left(\frac{\pi}{1-\pi}\right)^{\Delta^o + \hat{\theta}} - \frac{1-\chi}{\chi}\right] \tag{5}$$

where $\hat{\theta} = 2 \times \mathbb{1}\{\theta = H\} - 1$. Set $\Delta^o = \hat{\theta} = 0$ in (5). Note that the first multiplicand is positive, and it is multiplying a nonnegative term. This follows because $(1-\chi)/\chi \leq 1$ since $\chi \geq 1/2$. □

The next Lemma characterizes the contexts close to $1/2$ that lead to herding dynamics and the contexts close to 1 that lead all to pool on H.

Lemma 3. *Consider a game G with $\chi \geq 1/2$ and $\pi > (\beta+1)/2\beta$. Actor 1 uniquely pools on H if and only if the context χ is greater or equal to $\widehat{\chi} \in (\pi, 1)$.*

If $\pi < (\beta+1)/2\beta$ in game G, then there are no contexts above the cutoff $\widehat{\chi}$.

For both $x \in \{H, L\}$, an actor of type x believes most judges are of type x at the beginning of the game if and only if $\chi \in [1/2, \pi]$.

Proof. The value $\widehat{\chi}$ can be obtained by using Lemma 1:

$$P(\theta_2 = H \mid \theta_1 = L) = \frac{\beta+1}{2\beta} \tag{6}$$

$$\Leftrightarrow \left(\frac{\pi}{1-\pi}\right)^{-1}\left[\pi - \frac{\beta+1}{2\beta}\right] - \frac{1-\widehat{\chi}}{\widehat{\chi}}\left[\frac{\beta+1}{2\beta} - (1-\pi)\right] = 0$$

Solving for $\widehat{\chi}$ gives the result. That $\pi > (\beta+1)/2\beta$ holds if and only if $\widehat{\chi} < 1$ follows directly from the expression.

Suppose that the context is greater or equal to $\widehat{\chi}$. I now show that actor 1 uniquely pools on H.

Whatever the monotonic strategy prescription followed by actor 1, judges of type H judge actor 1 positively, and judges of type L judge actor 1 negatively. This is easy to see if actor 1 separates, or if he semi-pools on H and chooses L. It also follows if actor 1 chooses H and pools on H, since from (6) both types of judges would believe actor 1 is most likely of type H, and by Lemma 1 type H judges would believe he is of type H with probability at least $\pi^2 + (1-\pi)^2$. By the intuitive criterion and monotonicity of strategies, it also follows if actor 1 chooses L when he is pooling on H, since actor 1 of type H chooses H for sure. Finally, it also follows if actor 1 semi-pools on H, since judges would assign a higher posterior that actor 1 is of type H if he semi-pools on H and chooses H than if he pools on H and chooses H.

But then for any monotonic strategy, choosing H is preferred as long as the left-hand side of (6) is larger than the right-hand side, which is the condition that the context be above the cutoff $\widehat{\chi}$. Therefore, the unique MPBE strategy prescription is for actor 1 to pool on H.

To see that $\widehat{\chi} - \pi > 0$, note that the left-hand side is positive if and only if

$$\frac{\beta(2\pi-1)+1}{(1-\pi)(\beta(2\pi-1)-1)+\pi(\beta(2\pi-1)+1)} - 1 > 0 \Leftrightarrow$$

$$\beta(2\pi-1)+1-(1-\pi)(\beta(2\pi-1)-1)-\pi(\beta(2\pi-1)+1)$$

$$= (1-\pi)(\beta(2\pi-1)+1)-(1-\pi)(\beta(2\pi-1)-1) = 2(1-\pi) > 0, \text{ which always holds.}$$

Since $\chi \geq 1/2$, it must be that actor 1 of type H believes other individuals are most likely of type H at the beginning of the game. Consider a situation where, for both $x \in \{H, L\}$, an actor of type x believes most judges are of type x at the beginning of the game. This situation is characterized by actor 1 of type L believing other individuals are at least as likely to be of type L:

$$\frac{\pi(1-\pi)}{\chi - 2\chi\pi + \pi} \leq 1/2$$

$$\Leftrightarrow 2\pi^2 - (1+2\chi)\pi + \chi \leq 0$$

Solving for the equality of the above expression, we find that π is either equal to $1/2$ or χ. The left-hand side of the expression describes a parabola with a downward slope at $\pi = 1/2$ and an upward slope at $\pi = \chi$, therefore the inequality is satisfied for $\pi \leq \chi$. □

The next Lemma shows that as the group size grows, actors' expected proportion of judges and judges' probabilities over actors are arbitrarily close if both are of the same type duple.

Lemma 4. *Fix the amount of revealer judges and suppose there are only revealer and withholder judges. The difference between the probability assigned by a withholder judge of type θ to withholder actor i being of type θ, and actor i of type θ 's expected proportion of θ-types, or*

$$|P(\theta_i = \theta \mid (\theta_j^o, \theta_j) = w\theta, h_i) - \mathbb{E}_{j \neq i}(P(\theta_j = \theta \mid \theta_i = \theta, h_i))|,$$

tends to 0 as the group size increases.

Proof. Suppose e past actors have revealed they are type θ, and f past actors have revealed they are type $-\theta$. Then the expected proportion of judges of type θ for an actor i of type θ is:

$$\frac{e}{I} + \frac{I - e - f}{I} P(\theta_j = \theta \mid \theta_j^o = w, \theta_i = \theta, h_i) \tag{7}$$

The probability withholder judges of type θ place on withholder actor i being of type θ is the same as the probability actor i of type θ places on withholder judges being of type θ:

$$P(\theta_i = \theta \mid (\theta_j^o, \theta_j) = w\theta, h_i) = P(\theta_j = \theta \mid \theta_j^o = w, \theta_i = \theta, h_i)$$

□

The next Lemma shows that the threshold for herding on any action in a herding dynamic is finite for arbitrarily large groups. Note that this implies that the threshold for herding is finite in any group, since learning an actor's type provides more information about the distribution of preferences in a group the smaller that group is.

Lemma 5. *Consider a game G with $\chi \geq 1/2$, $\pi > (\beta+1)/2\beta$ and $I \to \infty$. The thresholds for uniquely pooling on θ in any period are finite in contexts below the cutoff $\widehat{\chi}$, and given by:*

$$\overline{N}^{max}(\theta; G) \equiv \tilde{\theta}(H) b_\infty + \left[\ln\left(\frac{\pi}{1-\pi}\right)\right]^{-1} \ln\left(\frac{1-\chi}{\chi}\right)$$

where $\tilde{\theta}(H) = (2 \times \mathbb{1}\{\theta = H\} - 1)$, and $b_\infty \equiv \dfrac{\ln(\beta(2\pi-1)+1) - \ln(\beta(2\pi-1)-1)}{\ln(\pi) - \ln(1-\pi)} + 1$.

If $\pi < (\beta+1)/2\beta$, the threshold for type θ to deviate from separating goes to ∞ as $I \to \infty$.

Proof. I will derive the threshold for uniquely pooling when $I \to \infty$. From Lemma 4, actor i of type H will uniquely pool on L if and only if he would deviate from separating:

$$1 + \beta \sum_{j \neq i} P(\theta_j = H \mid \theta_i = H, h_i) < \sum_{j \neq i} P(\theta_j = L \mid \theta_i = H, h_i)$$

$$\Leftrightarrow \sum_{j \neq i} P(\theta_j = L \mid \theta_i = H, h_i) \geq \frac{\beta+1}{2\beta}$$

The minimum information the actor needs to uniquely pool on L is given by setting the above expression as an equality. Using Lemma 1 and again using the fact that there are only pure MPBE strategies, we can rewrite the equality as follows:

$$\Bigg\{ P(\theta^o_{-i} = 1 \mid h_i, \theta_i = H) \left[1 - \frac{\beta+1}{2\beta}\right] - P(\theta^o_{-i} = -1 \mid h_i, \theta_i = H) \left[\frac{\beta+1}{2\beta}\right]$$
$$+ P(\theta^o_{-i} = 0 \mid h_i, \theta_i = H) \left[\pi - \frac{\beta+1}{2\beta}\right] \Bigg\} \left(\frac{\pi}{1-\pi}\right)^{\Delta(h_i)+1}$$
$$+ \Bigg\{ P(\theta^o_{-i} = 1 \mid h_i, \theta_i = H) \left[1 - \frac{\beta+1}{2\beta}\right] - P(\theta^o_{-i} = -1 \mid h_i, \theta_i = H) \left[\frac{\beta+1}{2\beta}\right]$$
$$+ P(\theta^o_{-i} = 0 \mid h_i, \theta_i = H) \left[1 - \pi - \frac{\beta+1}{2\beta}\right] \Bigg\} \left(\frac{1-\chi}{\chi}\right)$$
$$= \left(\frac{x_H}{I}\left[\frac{\beta-1}{2\beta}\right] - \frac{x_L}{I}\left[\frac{\beta+1}{2\beta}\right]\right)\left(\left(\frac{\pi}{1-\pi}\right)^{\Delta(h_i)+1} + \left(\frac{1-\chi}{\chi}\right)\right)$$
$$+ \frac{I - x_H - x_L}{I}\left[\left(\pi - \frac{\beta+1}{2\beta}\right)\left(\frac{\pi}{1-\pi}\right)^{\Delta(h_i)+1} + \left(\frac{\beta-1}{2\beta} - \pi\right)\left(\frac{1-\chi}{\chi}\right)\right] = 0$$

where x_H and x_L are the number of actors who have chosen H and L, respectively. Since $\Delta(h_i) = x_H - x_L$, it is generally hard to obtain a closed form solution of $\Delta(h_i)$ as a function of χ. However, this task is much simpler when $I \to \infty$, since the only term that remains is the term in square brackets. Notice that this term would always be negative if $\pi > (\beta+1)/2\beta$ does not hold. If $\pi > (\beta+1)/2\beta$ holds, and taking $I \to \infty$, through simple algebraic manipulations we get:

$$\Delta(h_i) = \frac{\ln(\beta(2\pi-1)-1) - \ln(\beta(2\pi-1)+1)}{\ln(\pi) - \ln(1-\pi)} - 1 + \left[\ln\left(\frac{\pi}{1-\pi}\right)\right]^{-1} \ln\left(\frac{1-\chi}{\chi}\right)$$

Analogous steps yield the amount of signals needed for actor i to uniquely pool on H:

$$\Delta(h_i) = \frac{\ln(\beta(2\pi-1)+1) - \ln(\beta(2\pi-1)-1)}{\ln(\pi) - \ln(1-\pi)} + 1 + \left[\ln\left(\frac{\pi}{1-\pi}\right)\right]^{-1} \ln\left(\frac{1-\chi}{\chi}\right)$$

□

The final Lemma shows that in a strategy profile where past strategy prescriptions have been pure, actors' and judges' beliefs becomes arbitrarily close as the group size grows. As opposed to Lemma 4, Lemma 6 considers any history that may arise in a pure strategy profile.

Lemma 6. *Consider a history where past actors have separated or pooled.*

For $x \geq 1/2$ and $y \in (x, 1)$, the following condition does not hold for a large enough group size:

$$P(\theta_i = \theta \mid (\theta_j^o, \theta) = w\theta, h_i) < x < y < \mathbb{E}_{j\neq i}(P(\theta_j = \theta \mid \theta_i = \theta, h_i)). \quad (8)$$

Proof. Take a history h_i in which past actors have separated or pooled.

Let z be the lead of disclosure $\hat{\theta}$.

Let $d/2$ be the number of revealers of type θ who can be matched with a revealer of type $-\theta$ in history h_i. The value d is the largest subset of revealers whose disclosure lead is equal to 0.

There are two cases to consider:

1) Actor i of type θ believes withholder judges are more likely to be type $-\hat{\theta}$: $P(\theta_j = \hat{\theta} \mid \theta_i = \theta, h_i, \theta_j^o = w) < 1/2$. The maximum expected proportion of actors of of type $\hat{\theta}$ comes from the

history h_i with the largest value of d. If $I - z$ is even, then this maximum proportion is equal to $z/I + .5(I-z)/I$. If $I-z$ is odd, then this maximum proportion is equal to $z/I + .5(I-z-1)/I + P(\theta_j = \theta' \mid \theta_i = \theta, h_i, \theta_j^o = w)$. By Lemma 5, there is a finite disclosure lead such that all actors pool on θ'. Therefore, z must be finite. But then, for I large enough, the maximum proportion is less than $y > 1/2$.

2) Actor i of type θ believes withholder judges are more likely of type $\hat{\theta}$: $P(\theta_j = \theta' \mid \theta_i = \theta, h_i, \theta_j^o = w) > 1/2$. For any period, the largest difference between the withholder judge's belief and the actors' belief comes from a history h_i where all past actors have revealed the same type (either $e = 0$ or $f = 0$ in (7) from the proof of Lemma 4).

Take the set of histories $\mathcal{H}'$ in which past actors have all revealed they are of type $\hat{\theta}$. Fix $x \geq 1/2$ and $y > x$, and suppose condition (8) does not hold for any history $h' \in \mathcal{H}'$. This is possible by Lemma 4, since for I large enough $P(\theta_i = \hat{\theta} \mid (\theta_j^o, \theta_j) = w\theta, h') < x$ implies $\mathbb{E}_{j \neq i} P(\theta_j = \hat{\theta} \mid \theta_i = \theta, h') < y$. Note there is a unique history h' for each lead of disclosure θ, which is equal to the lead of action θ.

Take a history h_i which has the same disclosure lead z as history $h' \in \mathcal{H}'$.

Notice that judges hold the same beliefs about withholder actor i in h_i and h' (Lemma 1):

$$P(\theta_i = \hat{\theta} \mid (\theta_j^o, \theta_j) = w\theta, h_i) = P(\theta_i = \hat{\theta} \mid (\theta_j^o, \theta_j) = w\theta, h')$$

In a history $h' \in \mathcal{H}'$, then it must be that either $\mathbb{E}_{j \neq i} P(\theta_j = \hat{\theta} \mid \theta_i = \theta, h_i) < y$ or $x < P(\theta_i = \hat{\theta} \mid (\theta_j^o, \theta_j) = wH, h_i)$.

Notice that the expected proportion of judges of type $\hat{\theta}$ is weakly lower in h_i than in h'. Indeed, the expected proportion of type θ' judges is equal to

$$\frac{z}{I} + .5\frac{d}{I} + \frac{I - z - d}{I} P(\theta_j = \hat{\theta} \mid \theta_i = \theta, h_i, \theta_j^o = w)$$

This term is decreasing in d.

But then $\mathbb{E}_{j \neq i} P(\theta_j = \hat{\theta} \mid \theta_i = \theta, h_i) < y$ or $x < P(\theta_i = \hat{\theta} \mid (\theta_j^o, \theta_j) = wH, h_i)$ hold for all histories in which past actors have pooled or separated, so condition (8) does not hold for any of these histories. □

B Main Proofs

In this section I collect the main proof of the results.

B.1 Proof of Proposition 1

Proof. By Lemma 3, actor 1 uniquely pools on H.

I now show that all actors uniquely pool on H. But this follows from the fact that if all actors $i' < i$ pool, actor i has the same incentives as past actors, since they have not revealed any information about the group. But then actor i uniquely pools, and this argument applies to all actors.

Let θ_I^m be the random variable representing the majority type of a group of size I. In contexts above the cutoff $\widehat{\chi}$, the MPBE probability of pluralistic ignorance is equal to

$$\chi P(\theta_I^m = L \mid \psi = H) + (1 - \chi)P(\theta_I^m = L \mid \psi = L) \tag{9}$$

First, note that since $P(\theta_I^m = L \mid \psi = H) < P(\theta_I^m = L \mid \psi = L)$, the MPBE probability of pluralistic ignorance diminishes with χ. Second, note that by the law of large numbers $P(\theta_I^m = L \mid \psi = H)$ diminishes with I and tends to 0, while $P(\theta_I^m = L \mid \psi = L)$ increases with I and tends to 1 (Casella and Berger, 2002). Therefore, the probability (9) goes to $1 - \chi$ as $I \to \infty$. The MPBE probability of pluralistic ignorance in a context above the cutoff $\widehat{\chi}$ can be made arbitrarily close to χ by making I arbitrarily large. The MPBE probability of pluralistic ignorance can then be made arbitrarily close to 0 by making I and χ large. □

B.2 Proof of Proposition 2

Proof. The proof contains several parts. First, I show the MPBE strategy prescriptions that arise for actors' and judges' possible beliefs. Second, I show how the incentives to separate and to uniquely pool change with the history of play. Third, I explicitly write the thresholds of a herding dynamic that arise in a MPBE. Fourth, I derive the upper and lower bounds of the probability of pluralistic ignorance for arbitrarily large groups. I also show that the MPBE probability of pluralistic ignorance with infinitely large groups depends on β, χ and π.

Part 1: MPBE strategy prescriptions that arise for actors' and judges' beliefs. Here I show the types of MPBE strategy prescriptions that arise as a function of actors' and judges' beliefs. I divide the beliefs that arise into two belief conditions, since by Lemma 2, an actor of type H believes more of his judges are of type H than does an actor of type L: $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \theta_i = H)\big)$ is larger than $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \theta_i = L)\big)$.

Belief condition 1. Each type of the actor believes most judges are of their own type. In this case, each type gets a higher expected judgment from revealing his type than from revealing the opposite type, or $\mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \theta_i = H)\big) < 1/2 < \mathbb{E}_{j\neq i}\big(P(\theta_j = H \mid h_i, \theta_i = H)\big)$, so separating is the unique prescription of a MPBE strategy.

Alternatively, both types may believe most judges are of type x.

Belief condition 2θ. Both types of the actor believe the majority type of the judges is type θ. There are two cases to consider. In order to characterize the MPBE strategy prescriptions that arise in each case, let q_i be the probability that actor i of type $-\theta$ chooses $-\theta$. Since type θ chooses θ (by monotonicity), $q_i \in [0, 1]$ fully describes all monotonic strategy prescriptions for actor i.

I will characterize the differential material payoff and the differential expected judgment as a function of q_i in each of the two cases. The *differential material payoff* is the difference in the material payoff to an actor of type $-\theta$ from choosing $-\theta$ minus choosing θ. Note that the differential material payoff is a constant in q_i, equal to 1. The *differential expected judgment* is the difference in the expected judgment to an actor of type $-\theta$ from choosing θ minus choosing $-\theta$.

For a given q_i, an actor of type $-\theta$ prefers to choose $-\theta$ if the differential material payoff is larger, and to choose θ if the differential expected judgment is larger.

Suppose i pools on θ (or $q_i = 0$). If i chooses $-\theta$, only type $-\theta$ judges will judge i positively. This uses the intuitive criterion to define out-of-equilibrium beliefs when the actor pools on θ: the surviving out-of-equilibrium belief would be that $-\theta$ deviated (by the monotonicity of strategy prescriptions). Without the intuitive criterion, we cannot rule out this out-of-equilibrium belief.

Case 1: Judges of type $-\theta$ judge i negatively if i pools on θ. Consider the differential expected judgment.

If actor i chooses $-\theta$, only judges of type $-\theta$ judge i positively.

If actor i chooses θ, judges of type $-\theta$ judge i negatively for any $q_i \in [0, 1]$. Judges of type duple (θ^o, θ) judge i negatively for a low enough q_i (or perhaps judge i positively for all q_i), and judge i positively for a large enough q_i. For some threshold value $\hat{q} \in [0, 1]$, judges of type duple (θ^o, θ) are indifferent in their judgment.

Therefore, the differential material payoff is an increasing step function of q_i.

The differential material payoff and the differential expected judgment are plotted in Figure IV. The Figure plots three differential material payoffs, labeled A, B and C. Separating is the unique MPBE strategy prescription for curve A, semi-pooling with $q_i = q_i^*$ is the unique MPBE strategy prescription for curve B, and pooling is the unique MPBE strategy prescription for curve C. It is easy to see from the Figure that a MPBE strategy prescription exists and is unique in this Case.

Note that if actor 1 is in belief condition 2θ, then he is in Case 1: all judges are withholders, and the probability actor 1 of type x assigns to judges being of type x' is the same probability judges of type x assign to actor 1 being of type x' at the beginning of the game. Further, since there are only withholder judges, the differential expected judgment curve is a straight line. So from imagining the squiggly curve as a flat line Figure IV, it can be seen that actor 1 either pools or separates. But then it follows that the probability of pluralistic ignorance is zero in sufficiently small groups.

By the last paragraph of Lemma 2, actor 1 is in belief condition 1 or in belief condition $2H$.

Note that actor 2 has a unique MPBE strategy to pool when $I = 2$. That is because actor 1 separates, and since actor 2 knows the type of his only judge, his differential expected judgment is

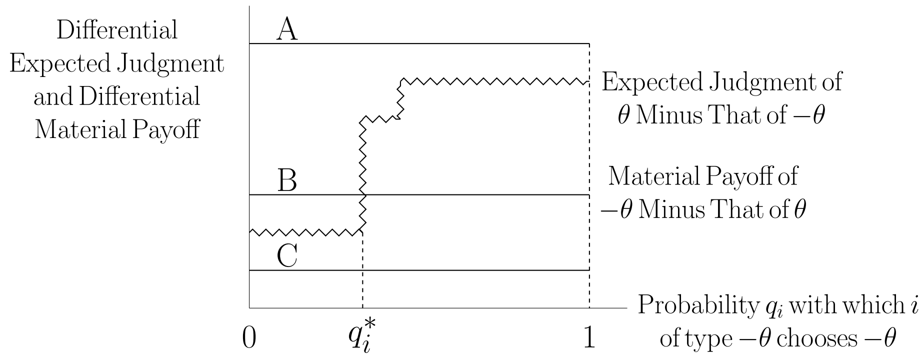

Figure IV: Uniqueness of the MPBE in Case 1 of Belief Condition 2θ. The graph depicts the differential expected judgment (squiggly curve) and the differential material payoff (straight lines) of an actor i of type $-\theta$ who believes most judges are of type θ, and who chooses $-\theta$ with probability q_i. Judges of type $-\theta$ judge i negatively for any q_i, while judges of type θ judge i positively for a large enough q_i. The graph depicts a situation in which separating is the unique equilibrium strategy prescription (denoted by the differential material payoff curve A), a situation in which mixing with $q_i = q_i^*$ is the unique MPBE strategy prescription (denoted by B), and a situation in which pooling on θ is the unique MPBE strategy prescription (denoted by C).

a step function as in Figure IV. But then there is uniqueness in sufficiently small groups.

Suppose that past actors have either separated or pooled. Then for I large enough, if actor i is in Case 1, all judges of type θ would judge him positively for choosing θ for all q_i. To see this, first note that for any $\varepsilon > 0$, withholder judges of type $-\theta$ believe that a withholder actor i is of type θ with probability larger than $1/2 - \varepsilon$ (Lemma 4). But revealer judges have one extra θ-signal than withholder judges of type $-\theta$. Then revealer judges believe that a withholder actor i is of type θ with probability larger than $\pi^2 + (1-\pi)^2 - \varepsilon$ (Lemma 1). But then revealer and withholder of type θ judges judge a withholder actor i positively. Since the probability that he is of type θ increases, they also judge him positively if he chooses θ for any q_i. But then the squiggly line in Figure IV would once again be flat, and actor i would have a unique pure MPBE strategy prescription.

Case 2: Some judges of type $-\theta$ judge actor i positively if i pools on $-\theta$. Consider the differential expected judgment. As before, if i chooses $-\theta$ only judges of type $-\theta$ judge actor i positively. Again as before, each judge of type θ either judges i positively for all q_i, or has a threshold $q \in [0, 1]$ such that they judge i positively for $q_i > q$, judge i negatively for $q_i < q$, and are indifferent in their judgment for $q_i = q$.

The difference with Case 1 is that now there are judges of type $-\theta$ who judge i positively for a low enough q_i, are indifferent in their judgment for a threshold value of q_i, and judge i negatively for a high enough q_i. The result is that the differential expected judgment curve may be non-monotonic in q_i.

An example is given in Figure V. The non-monotonicity of the differential expected judgment as a function of q_i implies that there may be multiple MPBE strategy prescriptions. In the Figure, $q_i = q'$, $q_i = q''$, and separating ($q_i = 1$) are all MPBE strategy prescriptions.

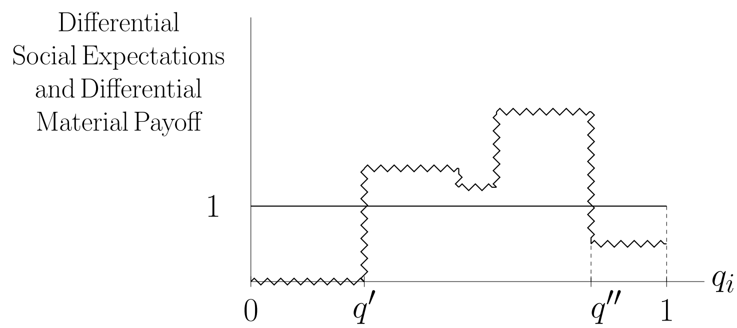

Figure V: Multiplicity of the MPBE strategy prescriptions in Case 2 of Belief Condition 2θ. The graph depicts the differential expected judgment (squiggly lines) and the differential material payoff (straight lines) of an actor i of type $-\theta$ who believes most judges are of type θ, and who chooses $-\theta$ with probability q_i. In the example, there are only withholder and revealer judges, and all judges assign a probability greater than $\overline{J}$ to i being of type $-\theta$. When i follows $q_i = 0$ and chooses θ, only judges of type $-\theta$ judge i positively. At q', withholder judges of type θ judge i positively as well. For a higher value of q_i, revealer judges of type $-\theta$ no longer judge i positively. For a higher value, revealer judges of type θ now judge i positively. For the larger value $q_i = q''$, withholder judges of type $-\theta$ no longer judge i positively. It is a MPBE strategy prescription to set $q_i = q'$, $q_i = q''$, and $q_i = 1$ (that is, to separate).

Since the differential expected judgment is convex-valued, a MPBE strategy prescription exists.

By Lemma 6, Case 2 does not exist for a sufficiently large I when past actors have separated or pooled. This follows from setting $x = 1/2$ and $y = \overline{J}$ in the statement of Lemma 6, and by noting that there would only be revealer and withholder judges.

Part 2: Incentives to separate and pool as the history of play changes. If actor i separates, only judges of type $-\theta$ would judge i positively for choosing $-\theta$, and only judges of type θ would judge i positively for choosing θ. The condition for i to not deviate from separating can therefore be written as $\mathbb{E}_{j\neq i}(P(\theta_j = \theta \mid h_i, \theta_i = -\theta)) \leq (\beta + 1)/2\beta$.

Consider two histories: h with corresponding actor i and h' with corresponding actor i', where i may be equal to i'. History h' is otherwise identical to h, except that h' leads the actor to assign a strictly higher probability to some judges being of type θ.

History h' leads to three differences in the actor's and judges' beliefs (by Lemma 2): (a) the actor i' assigns a higher expected probability to judges being of type θ, (b) the judgment of θ-types is weakly higher, and (c) the judgment of $-\theta$ types is weakly lower. By change (a), the incentives for actor i' of type $-\theta$ to deviate from separating are higher. By changes (a), (b) and (c), if the disclosure lead of θ is sufficiently high, actor i' uniquely pools on θ, even if i does not uniquely pool. Indeed, a history h' with an disclosure lead high enough leads to Case 1 of Belief condition 2θ, in which judges of type θ believe that if i pooled, he is of type θ with probability at least $(\beta + 1)/2\beta$, and actors believe there are at least $(\beta + 1)/2\beta$ of the type θ judges.

Part 3: The herding dynamics thresholds. Consider the action θ for which the disclosure lead is nonnegative at history h_i.

The difference $\underline{N}(\theta, h_i)$ minus the lead of θ is the size of the run of type θ needed for i of type $-\theta$ to deviate from separating. Similarly, $\overline{N}(\theta, h_i)$ minus the lead of action θ is the size of the run of type θ needed for i to uniquely pool on θ.

The lead of action θ minus the threshold $\underline{N}(-\theta, h_i)$ is the size of the run of type $-\theta$ needed for i of type θ to deviate from separating. The lead of action θ minus the threshold $\overline{N}(-\theta, h_i)$ is the size of the run of type $-\theta$ needed for i to uniquely pool.

Part 4: Bounds on the probability of pluralistic ignorance as group size grows. I now show that the upper and lower bounds for the MPBE probability of pluralistic ignorance when $\pi > (\beta + 1)/2\beta$ holds tend to $1 - \widehat{\chi}$ and $1 - \widehat{\chi}^{+}$ as the group size grows. An actor's action only affects others' belief about the realization of types through the information it reveals about the population and about the type of an individual in the group itself. The key observation is that, as the group size grows, an actor's type increasingly affects others' belief about the realization of types only through the information it reveals about the population. Indeed, any finite subset of the group becomes a negligible proportion of the group as group size increases.

As I grows there is an arbitrarily large proportion of states of the world in which actors herd. For sufficiently large I there is no Case 2 in Belief condition 2θ if past actors have separated or pooled (by Lemma 6, as mentioned in Part 1 of this proof). Suppose we are in Case 1. Take an actor of type $-\theta$ who deviates from separating, and consider his expected judgment from choosing θ when he is pooling. For sufficiently large I, these actors are judged positively by withholder judges of type θ and revealer judges (part 1 of this proof). The squiggly curve in Figure IV is therefore a flat line. Finally, the threshold for pooling is finite (by Lemma 5).

All actors uniquely pool on θ once there is enough information that the proportion of θ-types is large enough (by Part 2 of this proof). As I grows, the proportion is increasingly determined only by the updated population the group is drawn from, captured by the posterior of the context. In the limit, all actors uniquely pool on θ once actor i of type $-\theta$ assigns probability at least $\widehat{\chi}$ to the population being of type θ, and judges of type θ judge i positively for pooling.

The probability of pluralistic ignorance then tends to the probability that the majority type in the group is $-\theta$, which is given by $1 - \widehat{\chi}$. This establishes what the upper bound for the MPBE probability of pluralistic ignorance tends to as I increases.

To establish the lower bound, let us again apply our conclusion that as I grows, actors separate before herding in a MPBE for an arbitrarily high proportion of states of the world. But then the minimum probability of pluralistic ignorance that a herding dynamic can obtain in the limit is to incorporate one extra signal of type θ given a posterior context equal to the cutoff. Indeed, an actor will continue to separate when the posterior context is below $\widehat{\chi}$, and once he reveals he is of type θ, in the limit all actors uniquely pool on θ. Call this maximum posterior $\widehat{\chi}^{+}$, the formula for which is in Lemma 3. But then by the above logic, the lower bound on the MPBE probability of pluralistic ignorance is $1 - \widehat{\chi}^{+}$.

Computing the exact MPBE probability of pluralistic ignorance when a group is arbitrarily large is complicated, but we can establish that the value depends on β, χ and π. The value is equal to the multiplication of two terms which depend on θ, summed over θ. The first term is the probability of having a history that obtains a lead of action θ of size $\overline{N}^{max}(\theta, G)$ (Lemma 5) before a lead of action $-\theta$ of size $\overline{N}^{max}(-\theta, G)$. The second term is the probability of pluralistic ignorance given that the lead of action θ is $\overline{N}^{max}(\theta, G)$, and that all herd on θ. In particular, the second term is $1 - \chi'$, where χ' is the updated context given the lead of action θ of $\overline{N}^{max}(\theta, G)$. Since $\overline{N}^{max}(\theta, G)$ depends on β, χ and π, the MPBE probability of pluralistic ignorance when a group is arbitrarily large depends on β, χ and π (again, Lemma 5). □

B.3 Proof of Proposition 3

Proof. Much of the proof follows from Propositions 1 and 2.

For small enough groups, the probability of pluralistic ignorance is positive for contexts above the cutoff $\widehat{\chi}$ and 0 in contexts below the cutoff. But then the MPBE probability of pluralistic ignorance is lower when contexts are below the cutoff. Since the probability of pluralistic ignorance decreases with the context in the range of contexts greater or equal to $\widehat{\chi}$, the probabilities of pluralistic ignorance are higher in the subset $[\widehat{\chi}, \chi']$ than in any other context for all $\chi' \in (\widehat{\chi}, 1)$.

Conditional on the realizations of types in which actors chose pure MPBE strategy prescriptions, the probability of pluralistic ignorance in contexts below the cutoff $\widehat{\chi}$ has a lower bound and an upper bound that are arbitrarily close to $1 - \widehat{\chi}^{+}$ and $1 - \widehat{\chi}$, respectively. The MPBE probability of pluralistic ignorance for a context $\chi \in (\widehat{\chi}^{+}, 1]$ has an upper bound that tends to $1 - \widehat{\chi}^{+}$ as the group size increases (recall that $\widehat{\chi}^{+} > \widehat{\chi}$). Since the MPBE probability of pluralistic ignorance is decreasing in the context in the range of contexts greater or equal to $\widehat{\chi}$, then for any $\chi^{*} \in (\widehat{\chi}^{+}, 1]$ there exists a group size large enough that the MPBE probability of pluralistic ignorance is lower in the range of contexts $\chi \in [\chi^{*}, 1]$ than in any other context. The MPBE probability of pluralistic ignorance in context $\widehat{\chi}$ tends to $1 - \widehat{\chi}$ as the group size increases. But then, there is (generically) a range of contexts $[\widehat{\chi}, \chi^{**}]$ for some $\chi^{**} \in (\widehat{\chi}, \chi^{*})$ such that, for a group size large enough, the MPBE probability of pluralistic ignorance is higher in that range than in any other context. □

B.4 Proof of Corollary 1

Proof. The logic of both comparative statics can be aided by Figure IV. In sufficiently small or sufficiently large groups, the differential expected judgment is an increasing step function of q_i. This is shown for sufficiently large groups in part 1 of Proposition 2. It is also true in groups of size 2. Part 1 of Proposition 2 shows that it is always true for actor 1. But actor 2's only judge is judge 1, who at the beginning of period 2 has the same information than the information judge 2 had at the beginning of period 1.

Notice that since the differential expected judgment is multiplied by β, the step function is shifted upward with an increase in β. In sufficiently small groups, the probability of pluralistic ignorance is 0 for contexts below the cutoff $\widehat{\chi}$, so is not affected by β. The probability of pluralistic ignorance in contexts above the cutoff is $\chi(1-\pi)^2 + (1-\chi)\pi^2 > 0$. An increase in β does not affect this probability, but it strictly increases the range of contexts above the cutoff (it is easy to verify that $\partial\widehat{\chi}/\partial\beta < 0$). In sufficiently large groups, actors either pool or separate, and an increase in β leads an actor who separated at some history to uniquely pool, an actor who uniquely pooled on θ to continue uniquely pooling on θ. This weakly decreases the amount of information revealed by a group before actors herd. The probability of pluralistic ignorance therefore increases as well.

A similar upward shift in the differential expected judgment curve happens with an increase in the context χ when the actor is in belief condition $2H$. To see this, recall that the differential expected judgment curve is an increasing step function because only θ-type judges judge the actor positively for choosing θ, whatever his strategy prescription (from the proof of Proposition 2). But then, for any history, an increase in χ increases the probability that judges will judge him positively for choosing H, and decreases the probability that judges will judge him positively for choosing L. The upward shift in the differential expected judgment curve follows. Note that an increase in the context χ when the actor is in belief condition $2H$ leads to belief condition $2H$. In addition, the curve shifts to the left, since for any strategy prescription of the actor, judges assign a higher probability that the actor is of type H after the actor chooses H.

Through analogous reasoning to the above paragraph, an increase in the context χ when the actor is in belief condition $2L$ shifts the differential expected judgment curve downward and to the right. An increase in the context χ in belief condition $2L$ then either leads to belief condition $2L$, belief condition 1, or belief condition $2H$. But then the actor is less likely to choose L.

An increase in the context χ when the actor is in belief condition 1 either leads to belief condition 1, or to belief condition $2H$. The probability the actor chooses H therefore weakly increases.

But then the result follows. An increase in the context χ makes it weakly more likely that an actor chooses H in any history given any information actors have revealed.

In large groups, with an increase in χ actors then either change from separating to pooling on H, or from pooling on L to separating or to pooling on L, or do not change. In a group of size 2, the impact of the context χ on actor 1 is the same as for actors in large groups. Given the shifts in the differential expected judgment curve, the probability actor 2 chooses H is weakly higher after whatever information actor 2 reveals. □

Appendix References

Casella, George and Roger L Berger. 2002. *Statistical inference*. Vol. 2 Duxbury Pacific Grove, CA.

Online Appendices—For Online Publication

OA Minimal Example From Section 1 in Detail

This section provides the complete analysis of the minimal example from Section 1.

I work out the details of the PBE when there is no second-order uncertainty and high second-order uncertainty.

No second-order uncertainty. Individuals commonly know that their true attitudes were drawn from a population where 65% of individuals are pro-segregation. Alice then believes, no matter her true attitude, that Bob is pro-segregation with 65% probability.

If Alice's type-dependent strategy was to express her true attitude, she would have incentives to deviate if she were anti-segregation and cared enough about expected judgments. Indeed, Bob would be able to perfectly infer Alice's true attitude through her expressed attitude. But if Alice were anti-segregation and cared enough about expected judgments, she would deviate under the following condition:

$$\underbrace{1}_{\text{Util from expressing true attitude}} + \beta \times \underbrace{.35}_{\text{Expected judgment from expressing true attitude}} < 0 + \beta \times \underbrace{.65}_{\text{Expected judgment from expressing false attitude}} \Leftrightarrow \beta > 10/3$$

Alice would not deviate from a strategy in which she expressed a pro-segregation attitude no matter her true attitude, and expected judgments were high enough (that is, if β were greater than $10/3$). If Bob heard a pro-segregation attitude, he would believe Alice is pro-segregation with 65% probability. Since 65% is above the threshold for positive judgment (which is at most 54.5%), Bob would judge Alice positively if and only if Bob were pro-segregation, to which Alice assigns a probability of 65%. The intuitive criterion, a standard out-of-equilibrium refinement, implies that Bob would believe Alice is anti-segregation if she expressed an anti-segregation attitude. Alice would assign a 35% probability of being judged positively in that case. But then if Alice were anti-segregation and cared sufficiently about expected judgments, she would not want to deviate to express an anti-segregation attitude—again, $0 + \beta \times 0.65 > 1 + \beta \times 0.35 \Leftrightarrow \beta > 10/3$.

After Alice follows the strategy of expressing a pro-segregation attitude no matter her true attitude, Bob follows the same strategy. Indeed, Bob does not learn anything from Alice expressing a pro-segregation attitude, so he has the same incentives as Alice did in expressing his attitude. Subject to the out-of-equilibrium refinement, this is the unique equilibrium if $\beta > 10/3$ is low enough—otherwise, always expressing an anti-segregation attitude can also be sustained as an equilibrium strategy for either Alice or Bob.[15] However, only the latter equilibrium satisfies the monotonicity condition of Section 2.

Highest second-order uncertainty. Individuals are not sure which population their true attitudes were drawn from. They assign 50% probability to their true attitudes being drawn from a "mostly pro-segregation" population with 65% of pro-segregation individuals, and 50% to their true attitudes being drawn from a "mostly anti-segregation" population with 65% of anti-segregation individuals. These probabilities are common knowledge.

To form beliefs about Bob's true attitude at the beginning of the game, Alice uses her own true attitude as a signal of the population. A pro-segregation Alice will believe that the probability that the population is mostly pro-segregation is $(0.65 \times 0.5)/(0.65 \times 0.5 + 0.35 \times 0.65) = 0.65$, and an anti-segregation Alice will believe the probability is 0.35. If Alice is pro-segregation she will then believe that Bob is pro-segregation with probability $0.65 \times 0.65 + 0.35 \times 0.35 = 0.545$. By symmetry of the setup, if she were anti-segregation she would believe Bob is anti-segregation with probability 0.545.

Alice would not want to deviate from a strategy of expressing her true attitude. Bob would infer Alice's true attitude perfectly from her expressed attitude, so would judge Alice positively only if her expressed attitude equaled his true attitude. Then whatever her true attitude, she would believe that Bob would judge her positively with 54.5% probability. Alice therefore would not want

[15]Suppose that out-of-equilibrium beliefs are such that Alice would judge Bob positively if she were anti-segregation and would be indifferent in her judgment if she were pro-segregation. Since indifference may lead Alice to judge Bob positively with an arbitrarily high probability, Bob's expected judgments from choosing the out-of-equilibrium action could be set arbitrarily close to 1. Then, for these out-of-equilibrium beliefs, pro- or anti-segregation Bob would want to deviate from a strategy of always expressing an anti-segregation attitude if β is high enough. But then the intuitive criterion does not allow us to rule out either type, and out-of-equilibrium beliefs can sustain a strategy of always expressing an anti-segregation attitude—say, where pro-segregationists would judge actors positively for expressing an anti-segregation attitude.

to deviate from this strategy. Expressing her true attitude, whichever it is, gives her the greatest payoff:

$$1 + \beta \times 0.545 > 0 + \beta \times 0.455 \Leftrightarrow \beta > -1/0.09$$

For a β large enough, there also exist equilibria where Alice expresses the same attitude whatever her true attitude. Again, these alternative equilibria do not satisfy the monotonicity condition of Section 2.

After Alice expresses her true attitude, Bob will express whatever attitude Alice expressed, no matter his true attitude. Alice will judge Bob positively since Alice believes Bob is of her type with probability 0.545, which is above the threshold for judgment $\overline{J}$. When it is Bob's turn to express an attitude, he knows Alice's true attitude. Bob is more certain about Alice's true attitude than Alice was about Bob's in the case of no second-order uncertainty. But then we know that Bob will act analogously to how Alice acted with no second-order uncertainty and $\beta > 10/3$. The unique equilibrium outcome refined by the intuitive criterion is for Bob to express Alice's expressed attitude.

OB Working Out the MPBEs with Groups of Size 2, Arbitrarily Large Groups, and Some In Between Group Sizes

In this section I will consider equilibrium dynamics in a game G with groups of size 2, for infinitely large groups, and for some intermediate cases. I will show how equilibrium dynamics affect the MPBE probability of pluralistic ignorance differently in small and large groups. Section OB.1 shows through examples how the MPBE probability of pluralistic ignorance is non-monotonic in the context χ, the group size I, and the precision π. However, it also shows that the probability of pluralistic ignorance decreases monotonically with the contexts in the range of contexts above the cutoff $\widehat{\chi}$ when the group size is odd. Section OB.2 shows how, when there are multiple MPBEs, the strategy that minimizes the probability of pluralistic ignorance may include MPBE strategies which are not pure.

I assume throughout that $\overline{J} \to 1/2$.

OB.1 Non-Monotonicities of the MPBE Probability of Pluralistic Ignorance

Here I show that the MPBE probability of pluralistic ignorance is not monotonic in the context χ, the precision π, or the group size I. To do so, I provide a full analysis of groups of size 2 and of arbitrarily large groups. With these analyses I will show the non-monotonicities in χ and π, and show that the probability of pluralistic ignorance may increase in I when the context is below the cutoff $\widehat{\chi}$ and decrease in I when the context is above the cutoff. I then provide examples with groups of size 2, 3, 5 and 6 to show that the opposite may be the case. I conclude by showing that the MPBE probability of pluralistic ignorance decreases with the group size when the context is above the cutoff $\widehat{\chi}$ and we restrict attention to odd-numbered groups.

Group of size 2. Actor 1 always separates in contexts below the cutoff $\widehat{\chi}$, and uniquely pools in the contexts above the cutoff. The top graph of Figure VI summarizes the MPBE strategies as a function of the lead of action H and of the context. The bottom graph of the Figure plots the MPBE probability of pluralistic ignorance as a function of the context.

If the context is below the cutoff $\widehat{\chi}$, actor 2 pools on actor 1's choice unless actor 1 chose L and $\chi > \pi$. Actor 2 knows the type of individual 1, who is his judge, so I will call her judge 1. If $\chi < \pi$ and judge 1 does not learn anything about actor 2, she judges 2 positively (this can be seen from Lemma 1). Actor 2 will therefore uniquely pool on judge 1's type and be judged positively for sure. If $\chi > \pi$ and judge 1 does not learn anything about actor 2, she believes actor 2 is most likely of type H (again, by Lemma 1). If judge 1 is of type H, actor 2 uniquely pools on H and is judged positively for sure. If judge 1 is of type L, actor 2 of type H would deviate from revealing his type and from pooling on L—in both cases he is judged negatively for sure, and by deviating is either judged positively for sure or maximizes his material payoff. Actor 2 would uniquely semi-pool on L, making judge 1 indifferent in how she judges actor 2 after he chooses L. This example shows that an actor may uniquely semi-pooling for any β.

Regardless of the PBE strategy of actor 2, the MPBE probability of pluralistic ignorance is 0—actor 1 maximizes his material payoff, who makes up half of the group. However, the MPBE probability of pluralistic ignorance is positive in a context above the cutoff $\widehat{\chi}$. When both actors choose action H independent of their type, there is a probability $\chi(1-\pi)^2 + (1-\chi)\pi^2$ that both types are not choosing their type. This probability declines as χ increases.

Notice that *the MPBE probability of pluralistic ignorance is non-monotonic in* χ: it is equal to 0 for contexts below the cutoff $\widehat{\chi}$, reaches a maximum level for the cutoff, and is positive but decreasing for values of χ larger than the cutoff.

Similarly, *the MPBE probability of pluralistic ignorance is non-monotonic in* π. For $\pi = 1/2$, the MPBE probability of pluralistic ignorance is 0 since actor 1 separates—for any χ, he believes the second individual is equally likely to be of either type. If $\pi \in ((1+\beta)/2\beta, \chi)$ the context is above the cutoff, so the MPBE probability of pluralistic ignorance is positive. If $\pi > \chi$, the context is below the cutoff $\widehat{\chi}$, so the MPBE probability of pluralistic ignorance is 0.

Group of size ∞. Figure VII provides a graphical example of the PBE strategies and the probability of pluralistic ignorance for parameter values such that $\pi > (\beta+1)/2\beta$.[16] An individual's type provides information about the realization of types inasmuch as the type is informative about the population. Since the group is infinitely large however, an individual's type provides otherwise negligible information about the realization of types. For this reason, the thresholds for uniquely pooling are the same for all actors, and depend only on the lead $\Delta(h_i)$.

The lead needed to pool varies with the context. Consider the range of contexts from $1/2$ to x in the Figure. In that range, it takes a lead of size 2 to uniquely pool on either action. The MPBE probability of pluralistic ignorance is then $(1-\pi)^2/(1-2\pi(1-\pi))$,[17] which does not depend

[16]The parameter values corresponding to the Figure are $\beta = 1.9$, $\pi = 0.87$, $I \to \infty$, $x = 0.54$, $y = 0.86$, $z = 0.89$, $\widehat{\chi} = 0.98$, $\widehat{\chi}^{+} = 0.996$. The figure is not drawn to scale. By Lemma 5, actors separate for any χ if $\pi \leq (\beta+1)/2\beta$, and therefore there is no pluralistic ignorance. When $\pi > (\beta+1)/2\beta$ holds, however, actors will uniquely pool after some histories of play. Since I is infinitely large, this implies that there is a positive MPBE probability of pluralistic ignorance.

[17]The MPBE probability of pluralistic ignorance is the probability of herding on θ given that the population is $-\theta$ times the probability that the population is $-\theta$, summed over θ. The probability of herding on θ is the probability that there are two excess signals of type θ given a history in which there have not been two excess signals of type θ. The MPBE probability of pluralistic ignorance is then: $\sum_{x=0}^{\infty}[2\pi(1-\pi)]^x(1-\pi)^2(1-\chi) + \sum_{x=0}^{\infty}[2\pi(1-\pi)]^x(1-\pi)^2\chi$. This derivation is similar to the one found in Bikhchandani et al. (1992) for calculating the probability of herding on

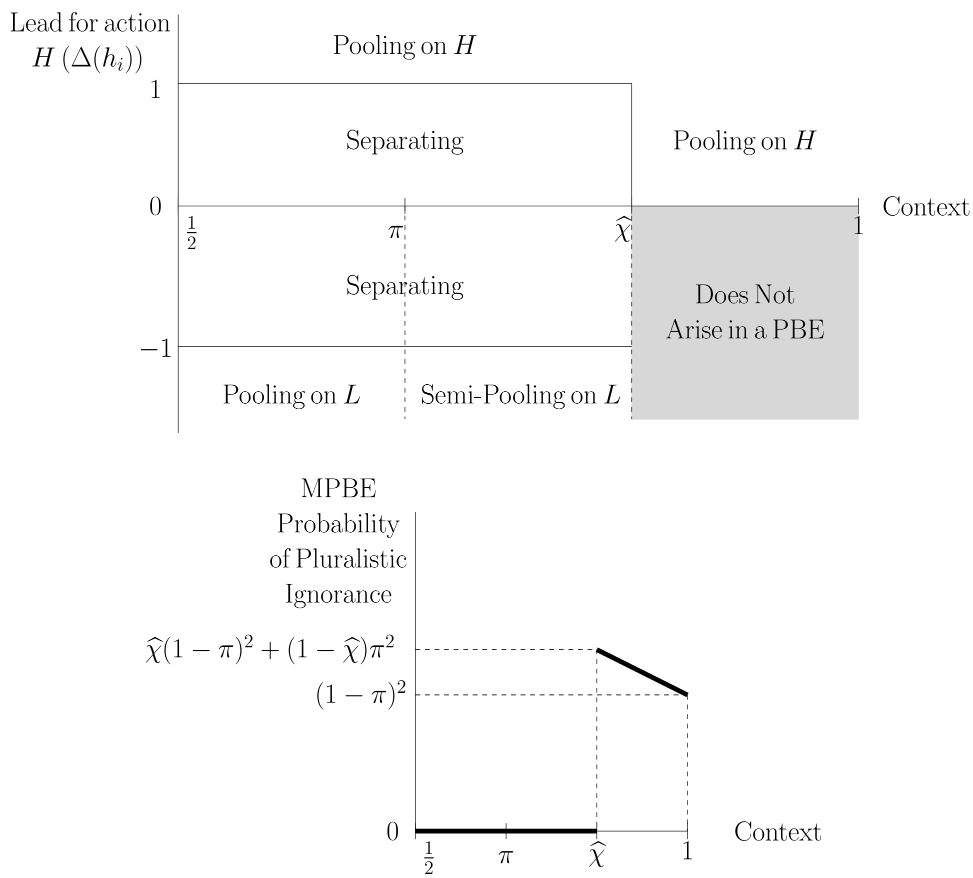

Figure VI: Group of Size 2. MPBE strategies (top) and the MPBE probability of pluralistic ignorance (bottom).

on χ given the symmetry of the leads needed to uniquely pool. For ranges of χ greater than x and lower than the cutoff $\widehat{\chi}$, it takes a lower lead to uniquely pool on H than to uniquely pool on L. For example, in the range of χ between x and y, the lead needed to uniquely pool is 2 for H, but 3 for L. Within this range of contexts, the MPBE probability of pluralistic ignorance declines with the context. This decrease comes from the fact that herding on H is more likely, so pluralistic ignorance is less likely the more likely the population is H.

The MPBE probability of pluralistic ignorance is non-monotonic in χ. For contexts between y and z, it takes an lead of 1 to uniquely pool on H, but a lead of 3 to uniquely pool on L. Actors then herd with less information about the group than in contexts between x and y, so it is strictly more likely that actors uniquely pool on the wrong population.

One way of seeing how much information is learned in a PBE is the difference between minimum information necessary for uniquely pooling and the information obtained in a PBE. Recall that the threshold for uniquely pooling on θ is $\overline{N}^{max}(\theta, h_i; G)$, which correspond to the two downward sloping lines on the top graph of Figure VII (Lemma 5). The thresholds capture the minimum information an actor needs to uniquely pool on θ. For a given χ, the gray area on the northern quadrant indicates how much extra information the group gets about the population being of type H before uniquely pooling on H. The gray area on the southern quadrant indicates how much extra information the group gets about the population being L before uniquely pooling on L.

Notice how the sum of these gray areas changes between the ranges $(1/2, x)$, (x, y), (y, z) and $(z, \widehat{\chi})$ in a way that corresponds to the MPBE probability of pluralistic ignorance on the bottom graph of Figure VII. This graphical representation helps illustrate that the MPBE probability of pluralistic ignorance will jump around in contexts below the cutoff $\widehat{\chi}$ when $I \to \infty$, but also that it is bounded (as per Proposition 2). It is bounded below by $1 - \widehat{\chi}$, the probability of pluralistic ignorance given the minimum information needed to uniquely pool. It is bounded above by $1 - \widehat{\chi}^{+}$,

the wrong state in their informational cascades model with two states and two actions. Indeed, when $I \to \infty$, expected judgment concerns incentivize matching the action to the population, analogous to how individuals want to match the action to the state in their informational cascades model. Note, however, that the extra certainty due to the precision $\pi > 1/2$ implies that the probability of pluralistic ignorance for a given χ is lower than the probability of herding on the wrong state for a signal precision equal to χ.

the probability of pluralistic ignorance given the maximum information an actor can receive in a PBE before he uniquely pools.

As before, the MPBE probability of pluralistic ignorance decreases with χ when $I \to \infty$ and the context is above the cutoff $\widehat{\chi}$. Recall that in these contexts all actors are choosing H. By the law of large numbers, the majority type of the group is equal to the majority type of the population for an infinitely large I. Therefore, for a context $\chi^{'}$ in the range of contexts above the cutoff, the MPBE probability of pluralistic ignorance is simply $1 - \chi^{'}$. But then the MPBE probability of pluralistic ignorance can be made arbitrarily small for a sufficiently high $\chi^{'}$. In particular, there is some $\chi^{''} \geq \widehat{\chi}^{+}$ in the range contexts above the cutoff such that the MPBE probability of pluralistic ignorance is lower for any $\chi \in [\chi^{''}, 1]$ than for any other context.

As illustrated by Figure VII, the MPBE probability of pluralistic ignorance may decrease with an increase in the context χ, as it may increase the lead needed to uniquely pool. It may also increase, establishing once again the non-monotonicity of the MPBE probability of pluralistic ignorance in χ.

Comparing groups of size 2 with an infinitely large group, the MPBE probability of pluralistic ignorance increases in contexts below the cutoff $\widehat{\chi}$ and decreases in contexts above the cutoff. However, we will now see that these changes are not monotonic.

Groups of size 5 and 6. Here I show that increasing the group size may decrease the MPBE probability of pluralistic ignorance in contexts below the cutoff $\widehat{\chi}$.

As a preliminary, note that if $\chi = 1/2$, $I \geq 3$ and $\beta \in (1, 2)$, actor 2 will always separate. First, note that if actor 1 reveals he is of type θ, then if actor 2 deviates from separating, it must be because actor 2 of type $-\theta$ deviates. Otherwise, actor 1 of type θ would have deviated from separating. Now, actor 2 of type $-\theta$ observes one θ signal and one $-\theta$ signal. Given the symmetry of the information structure, actor 2 of type $-\theta$ assigns a probability of $1/2$ to withholders being of type θ. So then consider the condition of actor 2 of type $-\theta$ for not deviating from separating:

$$1 + \beta \left[\frac{0}{I-1} + \frac{(I-2).5}{I-1} \right] > \beta \left[\frac{1}{I-1} + \frac{(I-2).5}{I-1} \right] \Leftrightarrow \beta < I - 1, \text{ with } I - 1 \geq 2$$

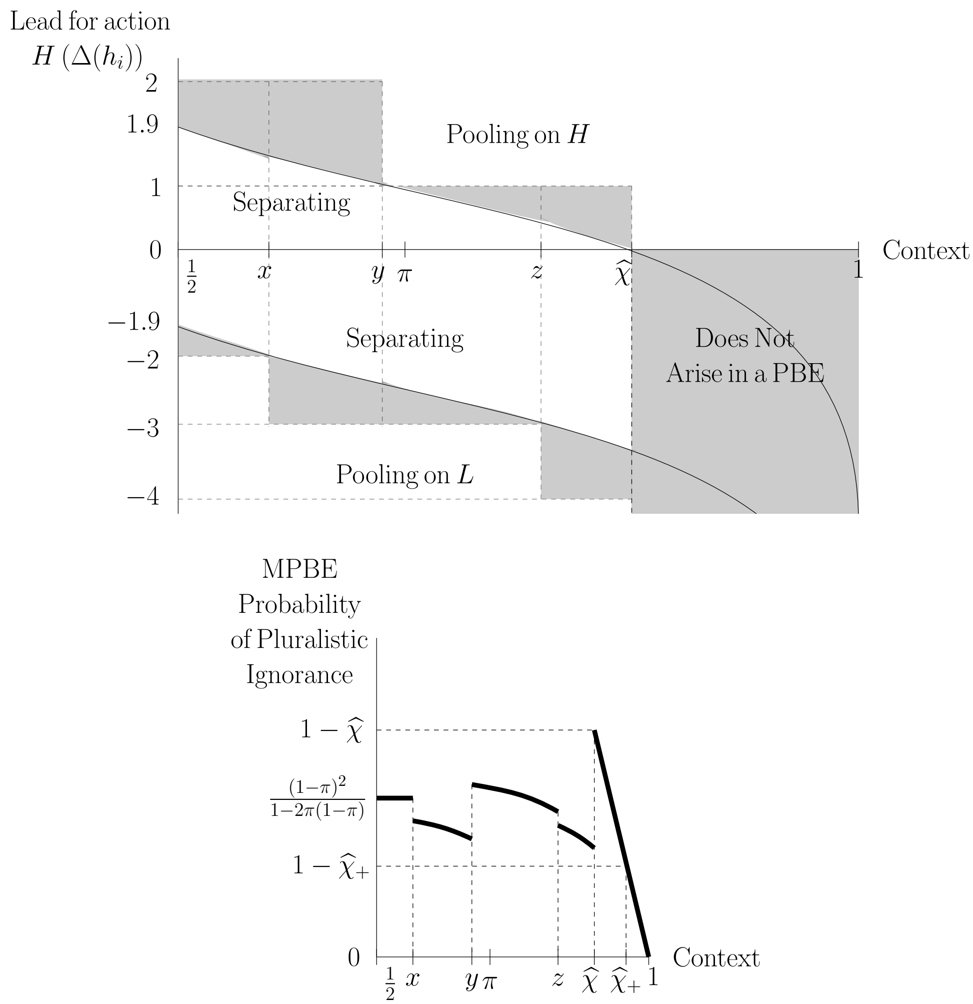

Figure VII: Group of Size ∞. MPBE strategies (top) and the MPBE probability of pluralistic ignorance (bottom).

where the left-hand side term is the expected proportion of H-type judges actor 3 of type L has in a groups of size 5, and the right-hand side term is the same for a group of size 6 (as per Lemma 1).

Since the MPBE probability of pluralistic ignorance is 0 for all $I \leq 4$, consider $I = 5$ and $I = 6$. We know the first two actors separate for $\chi = 1/2$. So consider the condition such that, after the first two actors reveal they are type H, actor 3 uniquely pools on H if $I = 5$, but separates if $I = 6$:

$$r(1)\frac{1+\pi}{2} + (1-r(1))\frac{1+(1-\pi)}{2} > \frac{\beta+1}{2\beta} > r(1)\left(\frac{2}{5}+\frac{3\pi}{5}\right) + (1-r(1))\left(\frac{2}{5}+\frac{3(1-\pi)}{5}\right)$$

This condition holds for parameter values $\beta = 1.9$, $\pi = 0.69$ and $\chi = 1/2$. Plugging these values into the above inequality yields 0.78 for the left term, 0.76 for the middle term and 0.74 for the right term. By symmetry of the information structure with $\chi = 1/2$, it must be that if the first two actors reveal they are type L, actor 3 uniquely pools on H if $I = 5$, but separates if $I = 6$.

There is no pluralistic ignorance if $I = 6$, since half the actors maximize their material payoffs. However, there is a positive MPBE probability of pluralistic ignorance if $I = 5$, since the first two actors' type may differ from the type of the last three individuals. With these parameter values, then, the MPBE probability of pluralistic ignorance decreases with an increase in I, since an increase in I increases the threshold for uniquely pooling.

But we can now conclude that *the MPBE probability of pluralistic ignorance is not monotonic in the group size I when the context is below the cutoff $\widehat{\chi}$.*

Groups of size 2 and 3. Here I show that the probability of pluralistic ignorance may increase with the group size in the range of contexts above the cutoff $\widehat{\chi}$.

In a context below the cutoff $\widehat{\chi}$, the MPBE probability of pluralistic ignorance is simply the probability that L is the majority type (by Proposition 1). In a group of size 2, this probability is given by $\chi(1-\pi)^2 + (1-\chi)\pi^2$. In a group of size 3, this probability is given by

$$\chi[(1-\pi)^4 + 4\pi(1-\pi)^3] + (1-\chi)[\pi^4 + 4(1-\pi)\pi^3]$$

It is easy to see that the MPBE probability of pluralistic ignorance is larger in a group of size 3 if the precision π is sufficiently close to $1/2$. Furthermore, the range of contexts above the cutoff $\widehat{\chi}$ may begin at a value of χ arbitrarily close to $1/2$ for β large enough. But then this establishes that *the MPBE probability of pluralistic ignorance is non-monotonic in the group size within the range of contexts above the cutoff.*

Monotonicity in odd-numbered groups with contexts above the cutoff $\hat{\chi}$. Despite the above non-monotonicity, I can establish that *the MPBE probability of pluralistic ignorance monotonically decreases in the range of contexts above the cutoff $\widehat{\chi}$ if we restrict the groups to be odd-numbered.*

Start out with all the states of the world of an odd-numbered group of size $I - 2$. The full description of a group of size I simply adds to each of those states of the world one of four options: two H-types, an H-type and then an L-type, an L-type and then an H-type, or two L-types. Since I is odd, the only situation in which the majority type passes from θ in the group of size $I - 2$ to $-\theta$ in the group of size I is when there is one more θ type than there are $-\theta$ types in a realization of types of the group of size $I - 2$, and we add two θ types to get a group of size I. There are an equivalent number of those states of the world when $\theta = H$ and when $\theta = L$. However, the states of the world in which $\theta = H$ are more likely to obtain since $\chi > 1/2$. But then there is an increase in the probability of a realization of types in which there is no pluralistic ignorance.

OB.2 The MPBE Probability of Pluralistic Ignorance May Be Minimized By Strategies Which Are Not Pure

I now provide an example in which there are multiple MPBEs, and the MPBE that minimizes the probability of pluralistic ignorance is not a pure-strategy MPBE.

Consider again a group of size 5 ($I = 5$), with weight on social expectations $\beta = 1.85$, precision $\pi = 0.9$ and context $\chi = 0.6$. With these parameter values, neither actor 1 nor 2 deviate from a separating strategy. The conditions for actor 3 to uniquely pool on θ if the first two actors revealed θ are:

$$r(1)\left(\frac{1}{2}+\frac{\pi}{2}\right)+(1-r(1))\left(\frac{1}{2}+\frac{(1-\pi)}{2}\right)>\frac{\beta+1}{2\beta}>\frac{\beta-1}{2\beta}>r(-1)\frac{\pi}{2}+(1-r(-1))\frac{1-\pi}{2}$$

The left-most term is the expected proportion of H types that an actor of type L has after observing two L signals (the expression follows from Lemma 1). The first inequality is the condition for actor 3 of type H to uniquely pool on L. Similarly, the right-most term is the expected proportion of L types that an actor of type H has after observing two H signals, and the third inequality is the condition of actor 3 of type L to uniquely pool on H.

With our parameter values, the left-most term is 0.92, the second term is 0.77, the third term is 0.22, and the right-most term is 0.1.

We are interested in semi-pooling strategies that may arise in a history of play that results in pluralistic ignorance. So suppose that actor 1 chose L. In order to sustain a semi-pooling PBE strategy prescription, it must be that actor 2 of type H would not deviate from a separating strategy, nor from a strategy of pooling on L. All judges would judge actor 2 positively if he was pooling on L, since all judges believe their own type is the majority type (this can be seen from Lemma 1). Only H type judges would judge actor 2 positively after a deviation (by monotonicity of strategy prescriptions and the intuitive criterion). In order for actor 2 to not deviate, it must be that actor 2 of type H believes judges are most likely of type L. Noting that $r(0)=\chi$, this condition reduces to:

$$\chi\frac{2\pi}{4}+(1-\chi)\frac{3(1-\pi)}{4}<\frac{1}{2} \tag{10}$$

To summarize, if actor 2 separated or pooled on L, actor 2 of type H believes most judges would judge him positively if he chose L. If actor 2 chose H for either of those strategy prescriptions, only H type judges would judge him positively (this uses the intuitive criterion, and monotonicity of strategy prescriptions). But then in order for a semi-pooling on L strategy to be sustained as a PBE strategy, it must be that actor 2 of type H does not deviate from pooling on L or from separating after actor 1 chose L:

$$\frac{\beta-1}{\beta} > \chi\frac{3\pi}{4} + (1-\chi)\frac{3(1-\pi)}{4} > \frac{\beta-1}{2\beta} \tag{11}$$

The first inequality ensures that the actor does not deviate from pooling on H, and the second inequality ensures that the actor does not deviate from separating.

Note that since $1/2 > (\beta-1)\beta$, the first inequality in (11) implies (10). This indeed holds with our parameter values, with the left-hand term equal to 0.45, the middle term equal to 0.43 and the right-hand term equal to 0.22. If (11) holds, then there is some randomization by withholder judges of type H that makes actor 2 of type H willing to randomize. The randomization of actor 2 of type H is determined by the condition that makes withholder judges of type H willing to randomize. The judges are indifferent in their judgment when, after seeing actor 2 choose H, they believe it is just as likely that he is of type H and type L:

$$P(\theta_2 = H \mid h_2, a_2 = L, \theta_j = H) = 1/2 \Leftrightarrow$$

$$\chi P(\theta_2 = H \mid h_2, a_2 = L, \psi = H) + (1-\chi)P(\theta_2 = H \mid h_2, a_2 = H, \psi = L) = 1/2$$

$$\Leftrightarrow \chi\left[\frac{q\pi}{q\pi+1-\pi}\right] + (1-\chi)\left[\frac{q(1-\pi)}{q(1-\pi)+\pi}\right] = \frac{1}{2}$$

$$\Leftrightarrow q^2 + q(2\chi-1)\frac{2\pi-1}{\pi-\pi^2} - 2 = 0$$

$$\Leftrightarrow q = \frac{1-2\chi}{2}\frac{2\pi-1}{\pi-\pi^2} \pm \sqrt{(2\chi-1)^2\left(\frac{2\pi-1}{\pi-\pi^2}\right)^2 + 8}$$

The step from the first to the second line follows since once you condition on the population, types are not informative about another individual's type. The step from the second to the third line follows because:

$$P(\theta_2 \mid h_2, a_2 = L, \psi) = \frac{P(\theta_2 = H, a_2 = L \mid h_2, \psi)}{P(a_2 = L \mid h_2, \psi)}$$

$$= \frac{P(a_2 = L \mid h_2, \theta_2 = H)P(\theta_2 = H \mid \psi)}{P(a_2 = L \mid h_2, \theta_2 = H)\pi + P(a_2 = L \mid h_2, \theta_2 = L)(1-\pi)}$$

where the first and second equalities use Bayes' rule, and the second inequality uses the fact that actor 2's strategy cannot condition on the population, while past history is irrelevant for determining individual 2's type once you condition on the population. The last line in the derivation gives a unique value of q, since the square root must have a positive sign for q to be positive.

With q in hand, we can use the definition of the observed type to calculate the observed type of actor 2 after choosing L:

$$\theta_2^o(h_2, a_2 = L) = \frac{\ln(q\pi + 1 - \pi) - \ln(q(1-\pi) + \pi)}{\ln \pi - \ln(1-\pi)}$$

Now we will check whether actor 3 has a PBE strategy to separate after observing actors 1 and 2 choose L, with actor 2 semi-pooling on L. The condition for actor 3 to separate is the following:

$$r(\theta_2^o)\left[\frac{q\pi}{q\pi + 1 - \pi} + \frac{2\pi}{4}\right] + (1 - r(\theta_2^o))\left[\frac{q(1-\pi)}{q(1-\pi) + \pi} + \frac{2(1-\pi)}{4}\right] > \frac{\beta - 1}{2\beta}$$

which again uses the derivation of $P(\theta_2 \mid h_2, a_2 = L, \psi)$ above. Indeed, with our parameter values this condition holds. In particular, $q = 0.78$, $\theta_2^o = -0.089$, the left-hand term of the inequality equal to 0.78, and the right-hand term of the inequality equal to 0.22.

But then we have found a PBE strategy with a lower probability of pluralistic ignorance than with a pure-strategy herding dynamic. In particular, with a pure-strategy herding dynamic there would be pluralistic ignorance if and only if the first two actors were of type θ and the rest were of type $-\theta$. In contrast, consider the following strategy profile. Actor 1 separates. Actor 2 separates if actor 1 chooses H, and semi-pools on L if actor 1 chooses L. Actor 3 uniquely pools on H if first and second actor choose H, and otherwise separates. Actor 4 uniquely pools on H if three of the earlier actors choose H, uniquely pools on L if the earlier actors all choose L, and separate otherwise. Actor 5 uniquely pools on H if three of the earlier actors choose H, uniquely pools on L if the earlier actors all choose L, has an undetermined strategy if the first three actors choose L and the third chooses H, and separates otherwise. The undetermined strategy is either to uniquely pool on L or to separate, is unique and is easy to determine, but does not affect the actions on the equilibrium path so I omit from the analysis. Given the analysis so far, this is indeed a PBE strategy for our parameter values. But then there will be pluralistic ignorance for sure if actors 1 and 2 are of type H and the last three are of type L, and there will be pluralistic ignorance with probability $q \in (0, 1)$ if the first and third actor are of type L, and the others are of type L.

For both PBE strategies, pluralistic ignorance happens only with runs that have the same probability of occurring. But unlike for the pure herding dynamic, for the PBE strategy these runs do not ensure pluralistic ignorance. *Therefore, the equilibrium probability of pluralistic ignorance may come from a PBE strategy profile that is not a pure-strategy herding dynamic.*

OC Easy-to-Compute Bounds on the Probability of Pluralistic Ignorance in Contexts Below the Cutoff $\widehat{\chi}$

In this section I provide some easy-to-compute bounds on the MPBE probability of pluralistic ignorance in contexts below the cutoff $\widehat{\chi}$. These bounds are particularly useful for intermediate-sized groups, since the main results provide the probability of pluralistic ignorance for small and large groups.

There are several challenges to calculating the upper and lower bounds in contexts below the cutoff $\widehat{\chi}$.

The first challenge is that, in a context below the cutoff $\widehat{\chi}$, the probability of of pluralistic ignorance is not monotonic in the group size I nor in the context χ (as shown in Section OB). This means that the extrema of the probability of pluralistic ignorance may not come from the same context χ for all group sizes. Furthermore, there is not a straightforward comparison of the MPBE probability of pluralistic ignorance between a context χ in which you need x excess H-signals to pool on H and y excess L-signals to pool on L, and a different context χ' in which you need $x' > x$ excess H-signals to pool on H and $y' < y$ excess L-signals to pool on L.

A second challenge is that the number of θ-signals needed to begin a herd on θ increases the more periods that have passed. The more periods that have passed without a herd, the more the current actor has observed a balanced number of H and L signals, which means that he needs to be very sure that the rest of the judges are not of his type in order to pool. Further, for a fixed x, the smaller the group or the more periods that have passed, the less likely that x excess θ-signals occur.

A third challenge is that actors pool on θ if a minimum of information about the distribution of types has been revealed, but actors may not herd until after more information has been revealed. This is a consequence of the fact that signals are discrete, so the minimum information may be surpassed before actors herd.

A fourth challenge is that, when actors semi-pool in equilibrium, calculating the probability of pluralistic ignorance becomes computationally more complicated. First, it requires calculating the probability with which an actor chooses an action (as in Section OB.2). Second, it requires keeping track of an increasing number of judges' type duples (see Section OD). Third, it opens the possibility that actors continue to semi-pool on some action θ without ever revealing enough information to herd on θ. Fourth, semi-pooling may lead actors to reveal more information about the distribution of types than would be the case if actors followed pure strategies (see Section OB.2).

In this section I provide upper and lower bounds on the probability of pluralistic ignorance that sidesteps all of these challenges. Let $\hat{\theta}$ be the majority type in the group. Both bounds will be a special case of $P_x \equiv P(\{\theta_k\}_{k=1}^x = H, \widehat{\theta} = L) + P(\{\theta_k\}_{k=1}^x = L, \widehat{\theta} = H)$, which equals

$$\begin{aligned}
&P(\widehat{\theta} = L \mid \{\theta_k\}_{k=1}^x = H, \psi = H)P(\{\theta_k\}_{k=1}^x = H \mid \psi = H)\chi \\
&\qquad + P(\widehat{\theta} = L \mid \{\theta_k\}_{k=1}^x = H, \psi = L)P(\{\theta_k\}_{k=1}^x = H \mid \psi = L)(1-\chi) \\
&\qquad + P(\widehat{\theta} = H \mid \{\theta_k\}_{k=1}^x = L, \psi = H)P(\{\theta_k\}_{k=1}^x = L \mid \psi = H)\chi \\
&\qquad + P(\widehat{\theta} = H \mid \{\theta_k\}_{k=1}^x = L, \psi = L)P(\{\theta_k\}_{k=1}^x = L \mid \psi = L)(1-\chi) \\
&= \pi^x \sum_{y \leq \lfloor \frac{I-1}{2} \rfloor - x} \binom{I-x}{y} \pi^y (1-\pi)^{I-x-y} + (1-\pi)^x \sum_{y \leq \lfloor \frac{I-1}{2} \rfloor - x} \binom{I-x}{y} \pi^{I-x-y}(1-\pi)^y
\end{aligned}$$

where the last step follows by symmetry of the information structure.

An Upper Bound on the Probability of Pluralistic Ignorance. In a context below the cutoff $\widehat{\chi}$, the first actor separates. To derive an upper bound, I will assume that all actors herd on actor 1's action. This clearly provides less information about the distribution of types than any MPBE.

Let $\hat{\theta}$ be the strict majority type in the group. Let $\overline{P}$ be the probability of pluralistic ignorance induced by a strategy profile in which actor 1 separates, all other actors pool on actor 1's type:

$$\overline{P}(I) \equiv \pi P(\hat{\theta} = \theta \mid \psi = -\theta, \theta_1 = \theta) + (1 - \pi) P(\hat{\theta} = \theta \mid \psi = \theta, \theta_1 = \theta)$$

A Lower Bound on the Probability of Pluralistic Ignorance. In order to avoid the complications that arise from semi-pooling, the lower bound will assume that if an actor semi-pools, then there is no pluralistic ignorance. We can use the Lemmas to show that if revealer judges believe that there is a probability of more than $\overline{J}$ probability that the population is θ, then actors separate or uniquely pool on θ if past actors have followed pure strategies and the observed type lead of θ is non-negative. The intuition is that if the assumptions hold, the material and expected payoff of type θ from choosing θ and $-\theta$ are the same whether the strategy prescription is to separate or to pool on θ (this uses the intuitive criterion). So I will focus on the probability of pluralistic ignorance when actors herd on L in histories in which the observed type lead of L has been non-negative in each earlier period.

To address the challenge of how the observed-type lead needed to herd changes with the group size and the period of play, one particularly conservative approach is to consider only the probability of pluralistic ignorance that arises when the first $\mathcal{N} \equiv \max_{\chi \geq 1/2, \theta \in \{H,L\}} \left| \lceil \overline{N}^{max}(\theta; G) \rceil \right|$ actors choose θ, where $|\lceil \cdot \rceil|$ is the absolute value of the ceiling function. Recall that $\overline{N}^{max}(\theta; G)$ is defined in Lemma 5 as the threshold for herding on θ in arbitrarily large groups, and is the largest initial run of θ needed to herd on θ. Also note that, from the formula in Lemma 5, the action θ and context $\chi \geq 1/2$ that maximize $\overline{N}^{max}(\theta; G)$ are $\theta = L$ and $\chi = \widehat{\chi}$. Therefore, $\mathcal{N} = \lceil -\overline{N}^{max}(L; \{\beta, \pi, \widehat{\chi}\}) \rceil$.

There is another reason to focus on action L. From the Lemmas, it can be shown that semi-pooling on H never arises in equilibrium, since the assumtpions $\chi \geq 1/2$ and $\overline{J}$ imply that if the observed type lead of type H is non-negative, an actor of type L only wants to deviate from separating when revealer and withholder judges of type H would judge him positively for choosing H when the actor pools on H.

The lower bound $\underline{P}(I)$ would then be $P_{\mathcal{N}}$ if the actors in the runs of size $\mathcal{N}$ we are considering choose a unique pure MPBE strategy prescription, and 0 otherwise. As per the logic of Proposition 2, for all groups above a certain threshold, there will be no semi-pooling strategies.

For odd-numbered groups, the probability of pluralistic ignorance increases monotonically in the lower bound. This follows from an analogous argument to the one made at the end of Section OB.1.

Although quite conservative, the lower bound at least provides an upper bound on the group size such that the probability of pluralistic ignorance is not 0. As the group size grows, the probability tends to $(1-\pi)^{\mathcal{N}}$.

Discussion. Both bounds can be made more precise by integrating more information regarding the thresholds for herding. The task retains a relative simplicity by focusing on histories of play in which actors follow pure strategy prescriptions.

Aside from their imprecision, a downside of the upper and lower bounds I have offered is that we know that the bounds are tighter for small groups and when the group size becomes arbitrarily large. Figure VIII provides an example which plots the upper and lower bounds I have proposed, as well as the precise bounds for arbitrarily large groups. The crosses $\times$ give possible values of the MPBE probability of pluralistic ignorance. If in groups of size 3 and 4 the first two actors separate, and in groups of size 5 and 6 the first three actors separate, the MPBE probability of pluralistic ignorance will be 0 for groups of size 2 to 6. However, the upper bound on the probability of pluralistic ignorance is only 0 for groups of size 2. As the group size grows, the proposed upper bound tends to a value that is higher than the upper bound from Proposition 2 (since $\widehat{\chi} > \pi$ by Lemma 3), and the proposed lower bound tends to a value that is higher than the lower bound from Proposition 2 (since in the lower bound I am only allowing pluralistic ignorance to happen in a subset of the realizations in which it would happen, and in which weakly more information is revealed than that corresponding to $\widehat{\chi}^{+}$).

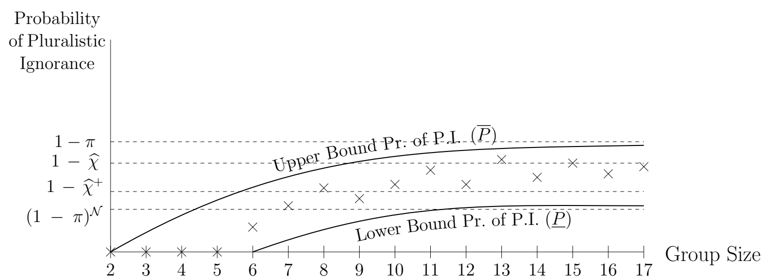

Figure VIII: Probability of pluralistic ignorance in a context below the cutoff $\widehat{\chi}$. The crosses ($\times$) give an example of the probability of pluralistic ignorance for some context below the cutoff $\widehat{\chi}$. The graph plots the easy-to-compute upper and lower bounds of the probability of pluralistic ignorance, as well as the upper and lower bounds of the probability of pluralistic ignorance in arbitrarily large groups according to Proposition 2.

OC.1 Sizing small and large groups

Corollary 2 uses the bounds to provide conservative estimates of the largest small group and largest small group. Here I provide a do-file for STATA that computes these bounds on group size.

Note that, when building the lower bound $\underline{P}$, the program ensures that there would be no semi-pooling for any $\overline{J}$ that satisfies the assumptions (again, that $\overline{J} \in (1/2, \pi^2 + (1-\pi^2))$.

```
//Here the user sets the parameter values of \pi, \beta and \chi


local pi=.7

local beta=3

local chi=.6


local chihat=(`pi'*(`beta'*(2*`pi'-1)+1))/
(`pi'*(`beta'*(2*`pi'-1)+1)+(1-`pi')*(`beta'*(2*`pi'-1)-1))
```

```
local smallgroupcondition=1
local largegroupcondition=1

// This part of the program calculates the largest small group,
//as long as the parameter values are consistent
// with the assumptions, and \chi is below
//the threshold \widehat{\chi}.

if `pi'>(`beta'+1)/(2*`beta') & `chi'<`chihat'{

while `smallgroupcondition'==1{
local I=`I'+1
local H=floor((`I'-1)/2)
local Bin1=binomial(`I',`H',`pi')
local PofIOne=`Bin1'
local BinOnePi=binomial(`I',`H'-1,`pi')
local BinOne1MinusPi=binomial(`I',`H'-1,1-`pi')
local OverlinePi=`pi'*`BinOnePi'+(1-`pi')*`BinOne1MinusPi'

if `PofIOne'<`OverlinePi'{
local smallgroupcondition=0
local J=`I'-1
}
}
display "Largest small group size is " `J'
}
```

```
else {
display "Condition \pi>(\beta+1)/(2*\beta) or condition
\chi<\widehat{\chi} do not hold"
}

local I=3
local chihat=(`pi'*(`beta'*(2*`pi'-1)+1))/(`pi'*(`beta'*(2*`pi'-1)+1)
+(1-`pi')*(`beta'*(2*`pi'-1)-1))
local smallgroupcondition=1
local largegroupcondition=1

// This part of the program calculates the smallest large  group.

if `pi'>(`beta'+1)/(2*`beta') & `chi'<`chihat'{

while `largegroupcondition'==1{
local I=`I'+2
local H=floor((`I'-1)/2)
local Bin1=binomial(`I',`H',`pi')
local Bin2=binomial(`I',`H',1-`pi')
local mathcalN=ceil(abs(-(ln(`beta'*(2*`pi'-1)+1)
-ln(`beta'*(2*`pi'-1)-1))/(ln(`pi')-ln(1-`pi'))-1
+(  ln((1-`chihat')/(`chihat')) / ln( (`pi')/(1-`pi') ) )))
local BinNPi=binomial(`I',`H'-`mathcalN',`pi')
local BinN1MinusPi=binomial(`I',`H'-`mathcalN',1-`pi')
local PofIOne=`Bin1'
local PofIChiHat=`chihat'*`Bin1'+(1-`chihat')*`Bin2'
```

```
local semipooling=0
local pitilde=`pi'/(1-`pi')
local chiundertilde=(1-`chi')/`chi'
local Jbar=`pi'^2+(1-`pi')^2

forvalues k=-1(-1)-`mathcalN' {
local r=(`pitilde'^(`k'))/(`pitilde'^(`k')+`chiundertilde')
local rrevealers=(`pitilde'^(`k'-1))/(`pitilde'^(`k'-1)+`chiundertilde')

if  `r'*((`I'-`k')/`I')*`pi'+(1-`r')*((`I'-`k')/`I')*(1-`pi')>
(`I'/(`I'-`k'))-(`beta'+1)/2*`beta'
& `r'*`pi'+(1-`r')*(1-`pi')<1-(`beta'+1)/2*`beta'
&  `rrevealers'*`pi'+(1-`rrevealers')*(1-`pi')<1/2
& `rrevealers'*`pi'+(1-`rrevealers')*(1-`pi')>1-`Jbar'  {

// The conditions `r'*((`I'-`k')/`I')*`pi'+(1-`r')*((`I'-`k')/`I')
//*(1-`pi')>(`I'/(`I'-`k'))-(`beta'+1)/2*`beta' &
//`r'*`pi'+(1-`r')*(1-`pi')<1-(`beta'+1)/2*`beta'
//imply that the actor deviates from pooling on L and from separating,
//respectively. I obtained it from using Lemma 1 to consider
//\mahbbm{E}_{j \neq i}(P(\mathcal{J}_{ji}=1 \mid
//h_i,\theta^o_i,\theta_i))>(\beta+1)/2\beta
//\Leftrightarrow (\#r/I)*0+((I-\#r)/I)*P(\theta_j=L \mid
//h_i,\theta^o_i=w,\theta_i)>(\beta+1)/2\beta
//\Leftrightarrow ((I-\#r)/I)*(1-P(\theta_j=L \mid
//h_i,\theta^o_i=w,\theta_i))>
//(\beta+1)/2\beta to derive the first condition,
```

```
//and an analogous reasoning for the second.

// The conditions ‘rrevealers’*‘pi’+(1-‘rrevealers’)*(1-‘pi’)<1/2 &
//‘rrevealers’*‘pi’+(1-‘rrevealers’)*(1-‘pi’)>1-‘Jbar’
//implies that the revealers’ beliefs are such that
//they judge the actor negatively if he pools.
//It is derived from Lemma 1 in a way analogous to the above paragraph.
//Note that ‘rrevealers’ has an observed type that is one less than ‘r’,
//since revealers have one more type $L$ signal as private
//information than do withholders of type $H$.

local semipooling=1
}
}

if ‘semipooling’==0{
local UnderlinePi=(‘pi’^‘mathcalN’)*‘BinNPi’
+((1-‘pi’)^‘mathcalN’)*‘BinN1MinusPi’
}

else{
local UnderlinePi=0
}

if ‘PofIOne’<‘UnderlinePi’ {
local largegroupcondition=0
local J=‘I’ + 2
```

```
}

}

display "The smallest odd-sized large group size is " `J'

}


else {

display "Condition \pi>(\beta+1)/(2*\beta) or condition
\chi<\widehat{\chi} do not hold"

}
```

OD Dynamics When Actors Semi-Pool on L

In this section I present the analysis of the equilibrium dynamics when actors semi-pool. The analysis is involved since there are a lot of moving parts. To simplify the exposition, the analysis will be presented graphically.

As a preliminary, I recall and define some concepts.

In Section A I define a judge's observed type, which is an important concept to understand for this section. Briefly, a judge j's observed type captures how much j's action revealed about his type. Suppose individual j semi-pooled on L, with type H choosing H with probability q_j. The larger q_j is, the higher the probability an actor $i > j$ places on judge j being of type L after observing that j chose L. Therefore, the larger q_j, the larger j's observed type if j chose L. I say j is a *semi-withholder* at history h_i if j semi-pooled on L and chose L by period i.

Semi-pooling occurs only when revealer judges believe actor i is more likely of type H than of type L, and type H actors believe judges on average are more likely of type L than of type H.

Consider first the case in which judges are either withholders or revealers, and in which at the beginning of period i all judges believe actor i is more likely to be of type H than type L. The left-hand side graph of Figure IX captures actor i's incentives. The horizontal axis denotes the probability q_i with which actor i of type H chooses H when semi-pooling on L. Notice that q_i uniquely identifies the prescription to semi-pool on L. The vertical axis captures differential payoffs, as I now explain.

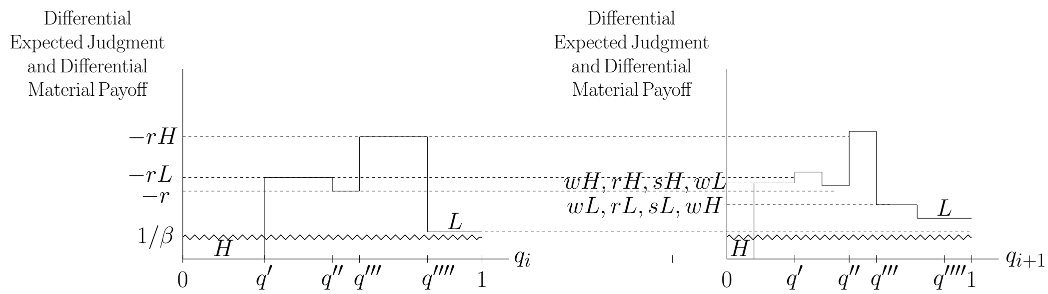

Figure IX: Left-Hand Side Graphs Depicts Semi-Pooling When There Are Only Withholder And Revealer Judges, And All Judges At The Beginning Of Period i Believe The Actor Is More Likely of Type H Than L. Right-Hand Side Graph Depicts Period $i+1$ After i Chooses L. The squiggly lines are the differential material payoff curve. The non-monotonic curves are the differential expected judgment curves.

The graph has two curves. The first is a flat line at $1/\beta$ (drawn squiggly in the picture). This line is the differential material payoff of actor i from choosing H instead of L (a single util), divided by the weight on expected judgment β. The second is a connected series of horizontal and vertical lines. This second curve captures the differential expected judgment from choosing L versus H. If actor i chooses H, he reveals he is of type H, and only judges of type H judge i positively. The variation in the differential expected judgment as q_i varies comes from the expected judges that will judge i positively from choosing L. The graph depicts which judges would judge i positively if i chose L. The Figure uses $\theta \in \{H, L\}$ to refer to θ-type judges, $r\theta$ to refer to revealer judges of type θ, $w\theta$ to refer to withholder judges of type θ, and $s\theta$ to refer to semi-withholder judges of type θ.

If $q_i = 0$, then all judges believe i is of type H, so only type H judges would judge i positively. As q_i increases, all judges' belief that i is of type L increases continuously (this can be seen from Lemmas 1 and 2). If $q_i = 1$, then all judges believe i is of type L, so only type L judges would judge i positively. There is then some value q' such that withholder judges of type L are indifferent in their judgment. This is represented by the horizontal part of the curve at q'. For some range of values of q_i larger than q', withholder judges of type L and judges of type H judge i positively, but not revealer judges of type L. This range ends at q'', where revealer judges of type H are indifferent in their judgment. There is then a range from q'' to q''' where all judges but revealers judge i positively. This range in which both revealers judge the actor positively appears due to the assumption that $\overline{J} > 1/2$. The rest of the curve is constructed via analogous reasoning.

For separating to not be prescribed by a PBE strategy, it must be that the curve of differential material payoff is lower than the curve of differential expected judgment at $q_i = 1$. Notice that in the left-hand side graph of Figure IX, since the differential material payoff is constant and the differential expected judgment is upper-hemi continuous and convex-valued, if i does not separate, semi-pooling on L is prescribed by a MPBE strategy in which withholder judges of type L are indifferent in their judgment. There are possibly other semi-pooling on L strategy prescriptions, in which revealer judges are indifferent and enough judge i negatively. This depends on how many revealer judges there are.

Suppose that i semi-pools on L and chooses L, and further suppose that at the beginning of period $i+1$ all judges believe actor $i+1$ is more likely of type H than of type L. This is feasible since it may take at least two L signals for judges to believe actors are more likely of type L than of type H. At period $i+1$ there are judges with six distinct type-duple combinations. There are withholders and revealers of type H and L, as before. But now there is also a semi-withholder of type H and a semi-withholder of type L. These semi-withholder judges correspond to actor i, whose observed type changed after his action.

The right-hand side of Figure IX shows the differential material payoff and the differential expected judgment of actor $i+1$. There are a few differences between the left-hand side and right-hand side of Figure IX.

The first difference is the presence of semi-withholders in the graph. Just to the left of the value of q_{i+1} such that revealers of type H are indifferent in their judgments, all judges except for revealers of type L judge $i+1$ positively, similar to before. However, there is a smaller value of q_{i+1} such that semi-withholders of type L are indifferent in their judgment. The value is q', the same value that made withholders of type L indifferent when i was the actor. Indeed, this is because i went from being a withholder in period i to a semi-withholder in period $i+1$. For values just to the left of q', only withholders of type L and H-type judges judge i positively. A similar logic applies to the differential expected judgment curve to the right of the value of q_{i+1} that makes revealer judges indifferent.

A second difference is that the curve shifts to the left. All judges except for i received information about the population that made them believe it is more likely that actor $i+1$ is of type L for every q_{i+1}. Therefore, the value of q_{i+1} that makes all judges except for i indifferent in their judgment is shifted to the left. This can be seen in the graph by comparing q' and q'' in the left-hand side and right-hand side graph.

The third difference is that $i+1$ believes that a judge who is not a revealer is more likely to be of type L than did i. This is reflected in the height of several parts of the differential expected judgment curve. Focus first on the left-hand side of Figure IX, which has labels on the vertical axis that represent the judges that judge i positively after choosing L. The differential expected judgment of choosing L instead of H is the expected proportion of judges on the vertical axis minus the expected proportion of H type judges. But then notice that the differential expected judgment for all $q_i > q'$ is the difference of a term that includes withholder judges of type L, minus a term that does not. That is, whenever it is not only withholder of type H judges that judge i positively for choosing L, the height of the left-hand side curve increases with judges' beliefs that the population is L. Therefore, compare the height of the differential expected judgment curve on both graphs of Figure IX. Whenever it is not only withholder of type H judges that judge i positively for choosing L, and the set of judges who judge the actor positively are the same, the height will be higher on the right-hand side graph.

As can be seen in the right-hand side of Figure IX, actor $i+1$ has a unique MPBE strategy prescription to semi-pool on L, in which $i+1$ of type H chooses H with a probability q_{i+1} smaller than the probability q_i with which i chose H. As was the case with actor i, $i+1$ may have other MPBE strategy prescriptions to semi-pool on L in which $i+1$ randomizes such that revealer judges are indifferent, and $i+1$ is indifferent if the revealer judges are sufficiently likely to judge $i+1$ negatively. Again, there would have to be enough revealers for this semi-pooling to be prescribed by a PBE strategy.

More generally, that semi-pooling on L may be prescribed by a MPBE strategy follows from the fact that the differential material payoff curve is a horizontal line and the differential expected judgment curve is upper-hemi continuous and convex-valued. Therefore, if the differential material payoff is lower than the highest differential expected judgment, the curves cross at some point. Otherwise, the PBE strategy prescribes the actor to separate.

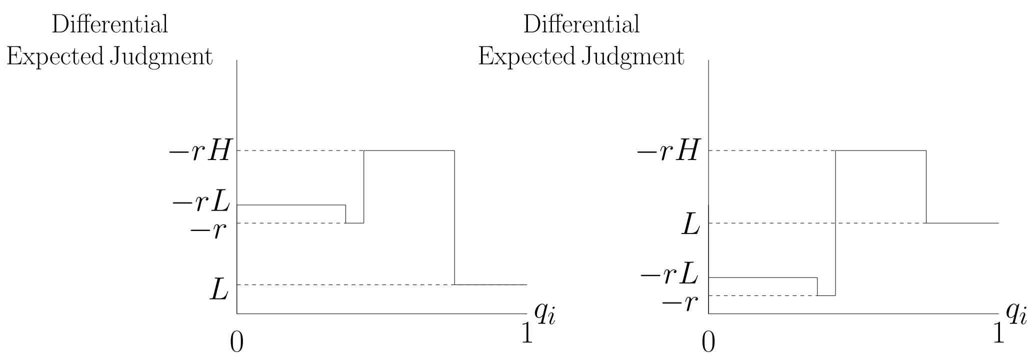

Figure X: Differential Expected Judgment When Withholder Judges Of Type L Judge i Positively If He Pools

There is a second condition under which revealer judges believe actor i is more likely of type H than of type L, and type H actors believe judges on average are more likely of type L than of type H. This is the condition that withholder judges of type L would judge i positively if i pooled. The differential expected judgment of this case are depicted in Figure X for the case where no actors have semi-pooled. Much of the analysis of this case is similar to that of the case we just analyzed.

What is different about the case depicted in Figure X is that pooling on L may be prescribed by the MPBE strategy. This would happen if $1/\beta$ is below the differential expected judgment curve when $q = 0$. Notice, moreover, that both a pooling on L and semi-pooling on L may be prescribed by a PBE strategy. This can be seen in the right-hand side graph, for a value of $1/\beta$ between the $-rL$ judges and the $-r$ judges.

Online Appendix References

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