Pluralistic Ignorance and Group Size∗
A Model of Expressing Attitudes with Image Concerns

Mauricio Fernández Duque†

February 12, 2019

Abstract

I develop a theory of group interaction in which individuals who act sequentially are concerned with signaling what they believe is the majority type. Group members judge an individual positively if they believe her type matches their own. Individuals distort their behavior to be judged more positively. Social learning may result in most individuals acting according to what they mistake for the majority type, a situation known as pluralistic ignorance. Two key determinants of pluralistic ignorance are the group size and the ex ante information individuals have about its composition. In particular, individuals are uncertain over which distribution types were drawn from. Pluralistic ignorance is minimized in small groups when this uncertainty is high, and minimized in large groups when it is low. When pluralistic ignorance occurs, actions are inefficient: individuals pander to a mistaken sense of the majority view and, perversely, earn the majority’s disapproval. I consider the optimal policy of a benevolent social planner with private information about the distribution of types. The planner reveals her information in large groups, but makes the minority appear larger in small groups.

Keywords: Pluralistic ignorance, group size, image concerns, expected judgment, social learning

JEL Codes: C72, D83, D90

Word count: 23,121 for the main text, 6,891 for the supplementary material.

∗I would like to thank Samuel Asher, Nava Ashraf, Robert Bates, Iris Bohnet, Filipe Campante, Paul Feldman, Jeffry Frieden, Benjamin Golub, Michael Hiscox, Antonio Jiménez, Horacio Larreguy, Navin Kartik, Scott Kominers, Stephen Morris, Carlo Prato, Pia Raffler, Matthew Rabin, Kenneth Shepsle, Joel Sobel, and seminar participants at Harvard University, ITAM, CIDE, and the EEA-ESEM 2018 meeting.

†CIDE. Carretera México-Toluca 3655, Col. Lomas de Santa Fe, 01210, Mexico City, Mexico. E-mail: mauricio.fernandez@cide.edu.
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 its support, 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).\(^1\) Whites’ public opinion regarding segregation is an example of pluralistic ignorance at a large scale. Pluralistic ignorance is also found in 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).

This paper studies how pluralistic ignorance arises when individuals have image concerns that are shaped by social learning.\(^2\) 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-segregationist attitude, but may express either attitude. Others in the group judge those who have expressed an attitude. Alice judges Bob positively if and only if Alice believes Bob’s true attitude is more likely than not equal to her own true attitude.\(^3\) 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

---

\(^1\)Pluralistic ignorance has been found in a variety of public opinion issues, including religious observance (Schanck, 1932), climate change beliefs (Mildenberger and Tingley, 2017), tax compliance (Wenzel, 2005), political correctness (Van Boven, 2000), unethical behavior (Halbesleben, Buckley and Sauer, 2004), pre-election Trump support Bursztyn, Egorov and Fiorin (2017), support for female labor force participation in Saudi Arabia (Bursztyn, González and Yanagizawa-Drott, 2018), and alcohol consumption (Prentice and Miller, 1993). There is a large empirical literature on public opinion (Noelle-Neumann, 1974, Katz and Allport, 1931, Shamir and Shamir, 2000, J O’Gorman, 1986). There is also a large theoretical literature on equilibrium misaggregation of information with both Bayesian agents (Acemoglu et al., 2011, Bikhchandani, Hirshleifer and Welch, 1992, Banerjee, 1992, Smith and Sørensen, 2000, Goeree, Palfrey and Rogers, 2006, Lee, 1993), and non-Bayesian agents (Eyster and Rabin, 2010, Golub and Jackson, 2010, Bohren, 2016) – see Golub and Sadler (2016) for a review.

\(^2\)The paper is related to the literature on conformity (e.g. Bernheim, 1994, Ellingsen, Johannesson et al., 2008, Benabou and Tirole, 2012) and social image concerns (see Bursztyn and Jensen, 2017, for a review).

\(^3\)For instance, because a positive judgment indicates a willingness to interact in the future, and Alice is biased towards those that have similar true attitudes as her. For a similar formulation, see Bursztyn, Egorov and Fiorin (2017).
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, and I will study when these three features appear in equilibrium with a positive probability. First, most group members **reluctantly** express an attitude, in the sense that the attitude they express differs from their true attitude – many whites expressed support for segregation even though they did not really support it.\(^4\) Second, group members underestimate others’ reluctance – whites believed that those who expressed support for segregation did really support it. Third, underestimating others’ reluctance makes individuals more willing to reluctantly express an attitude – whites were more willing to express support for segregationist housing policies when they overestimated others’ support.\(^5\)

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 his prior 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. In particular, true attitudes are drawn from one of two populations

---

\(^4\)I take the term ‘reluctance’ from Dana, Cain and Dawes (2006), a seminal experiment showing that individuals act according to what they think others expect, and would act differently in the absence of these expectations. Also see Cain, Dana and Newman (2014).

\(^5\)Rational explanations of pluralistic ignorance can be found in Kuran (1997), Bicchieri (2005), Centola, Willer and Macy (2005), Chwe (2013), while psychological explanations are reviewed in Kitts (2003) and sociological explanations in Shamir and Shamir (2000).
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. Indeed, one’s own type provides a signal of the population from which they are drawn, a signal that is very informative with high second-order uncertainty. First movers will express their true attitude, believing most will judge them positively, and in so doing they 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.

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 probability of pluralistic ignorance is lowest in large groups when second-order uncertainty is low, but lowest in small groups when second-order uncertainty is high. 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 an increasingly large fraction of the group as group size diminishes. In the next section I provide a minimal example with which I elaborate on the logic of the main result.

To study problems of cooperation, I extend the model so that an expressed attitude generates a positive externality towards group members with the same true attitude. Under this interpretation, ‘expressing an attitude’ can include protesting, providing a decentralized public good, gift giving and engaging in a joint project. I show that the main result continues to hold in this more general framework. Moreover, by considering externalities, we can consider social welfare and policy implications.

Pluralistic ignorance may lead to a perverse failure of collective action: most individuals avoid
expressing their true attitude, instead expressing the one they mistakenly believe increases social welfare. This results in frayed social interactions, since most individuals judge most others negatively, mistakenly believing they do not share their true attitude.\(^6\)

A social planner can avoid this perverse failure of collective action if she has private information about the population from which the group’s attitudes were drawn and can commit \textit{ex ante} to an announcement strategy. The main result implies that a social planner should decrease second-order uncertainty in large groups and increase it in small groups. A social planner facing a large group, such as a policymaker trying to avoid pluralistic ignorance in a public opinion topic, should then reveal all her information about the population.\(^7\) A social planner facing small groups, such as the Dean of a university who wants individuals to express minority attitudes in conversations, should then increase the perceived size of the minority.

When the externalities of expressing an attitude are particularly high, pluralistic ignorance may be socially optimal. A Dean may want students to believe that most students are not racists even if most actually are, because such a belief would diminish acts of violence. In this case, the Dean would want to disseminate information to increase the probability that most will express an anti-racist attitude, whether or not it creates pluralistic ignorance.

Section 1 shows the logic of the main result with a minimal example. In section 2 I set up the general model; I present the main results in sections 3 and 4. Section 3 describes equilibrium dynamics, and section 4 shows how pluralistic ignorance varies with group size. In section 5 I extend the basic setup to allow actions to have externalities. Section 6 considers the social welfare consequences of pluralistic ignorance with and without externalities. Section 7 studies the optimal information dissemination policy of a social planner who has private information about the population, and how the policy varies with group size. Section 8 discusses how to empirically test the model. Section 9 provides a review of the theoretical and experimental literature on pluralistic

---

\(^6\)This perversity relies on a particular interpretation of utility functions and of the social welfare function, a point I develop in section 6. The perverse failure of collective action provides a link between the literature on public opinion with the largely independent literature on collective action problems (Olson, 1965, Hardin, 1982, Esteban, 2001, Siegel, 2009).

\(^7\)There is a growing literature on campaigns that disseminate information about the distribution of attitudes, as well as its relationship to pluralistic ignorance, which I discuss in section 9.
ignorance. Section 10 concludes.

1 A Minimal Example

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-segregationist ($\theta_i = H$) or anti-segregationist ($\theta_i = L$) – this is their true attitude. The conversation is highly stylized: first, Alice expresses a pro-segregationist attitude or an anti-segregationist attitude, and then Bob expresses a pro-segregationist attitude or an anti-segregationist attitude. The individuals get one util from expressing their true attitude, and zero util otherwise. In addition, they put weight $\beta > 1$ on how they think they will be judged by the other individual – I call this expected judgment their social expectation. They care about being judged positively (perhaps due to reputational concerns), and trade that off against expressing their true attitude. An individual judges the other positively if the other individual’s type is most likely equal to her own type. Thus, Alice’s social expectation is:

$$P(P(\theta_A = \theta \mid \theta_B = \theta) > 1/2 \mid \theta_A),$$

or the probability that Alice places on Bob believing she is most likely of the same type. Notice the importance of second-order beliefs when forming social expectations.

Bob’s social expectation is similar to Alice’s, but before forming his beliefs Bob observes the expressed attitude $a_A$ of Alice:

$$P(P(\theta_B = \theta \mid \theta_A = \theta) > 1/2 \mid \theta_B, a_A).$$

How does this conversation play out? It depends on the information the individuals have of each other at the outset – in particular, on their second-order uncertainty. I will consider two extreme examples of this uncertainty.
No second-order uncertainty. Individuals commonly know that their true attitudes were drawn from a population where 80% of individuals are pro-segregationist. Alice then believes, no matter her true attitude, that Bob is pro-segregationist with 80% probability.

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

$$1 + \beta \times .2 < 0 + \beta \times .8 \iff \beta > \frac{5}{3}$$

Util from expressing true attitude + Social expectation from expressing true attitude < Social expectation from expressing false attitude ⇔ \( \beta > \frac{5}{3} \)

Alice would not deviate from a strategy in which she expressed a pro-segregationist attitude no matter her true attitude, and social expectations were high enough (that is, if \( \beta \) were greater than \( \frac{5}{3} \)). If Bob heard a pro-segregationist attitude, he would believe Alice is pro-segregationist with 80% probability. Bob would judge Alice positively if and only if Bob were pro-segregationist, to which Alice assigns a probability of 80%. The intuitive criterion, a standard out-of-equilibrium refinement, implies that Bob would believe Alice is anti-segregationist if she expressed an anti-segregationist attitude. Alice would assign a 20% probability of being judged positively in that case. But then if Alice were anti-segregationist and cared sufficiently about social expectations, she would not want to deviate to express an anti-segregationist attitude – again, \( 0 + \beta \times .8 > 1 + \beta \times .2 \iff \beta > \frac{5}{3} \).

After Alice follows the strategy of expressing a pro-segregationist attitude no matter her true attitude, Bob follows the same strategy. Indeed, Bob does not learn anything from Alice expressing a pro-segregationist 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 \( \beta > \frac{5}{3} \) is low enough – otherwise, always expressing an anti-segregationist attitude can also be sustained as an
equilibrium strategy for either Alice or Bob.\footnote{Suppose that out-of-equilibrium beliefs are such that Alice would judge Bob positively if she were anti-segregationist and would be indifferent in her judgment if she were pro-segregationist. Since indifference may lead Alice to judge Bob ‘arbitrarily positively’ (in a sense I specify in section 2), Bob’s social expectations from choosing the out-of-equilibrium action could be set arbitrarily close to one. Then, for these out-of-equilibrium belief, anti-segregationist Bob would want to deviate from a strategy of always expressing a pro-segregationist attitude if $\beta$ is high enough. But then the intuitive criterion does not hold, and out-of-equilibrium beliefs can sustain a strategy of always expressing an anti-segregationist attitude.}

_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-segregationist’ population with 80% of pro-segregationist individuals, and 50% to their true attitudes being drawn from a ‘mostly anti-segregationist’ population with 80% of anti-segregationist 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-segregationist Alice will believe that the probability that the population is mostly pro-segregationist is $(0.8 \times 0.5)/(0.8 \times 0.5 + 0.2 \times 0.5) = 0.8$, and an anti-segregationist Alice will believe the probability is 0.2. If Alice is pro-segregationist she will then believe that Bob is pro-segregationist with probability $0.8 \times 0.8 + 0.2 \times 0.2 = 0.68$. By symmetry of the setup, if she were anti-segregationist she would believe Bob is anti-segregationist with probability 0.68.

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 68% probability. Alice therefore would not want to deviate from this strategy. Expressing her true attitude, whichever it is, gives her the greatest payoff:

$$1 + \beta \times 0.68 > 0 + \beta \times 0.32 \Leftrightarrow \beta > -1/0.36$$

For a $\beta$ large enough, there also exist equilibria where Alice expresses the same attitude whatever her true attitude. Given that I am interested in information misaggregation, I am making a conservative assumption in selecting the equilibrium strategy that reveals most of Alice’s information.
After Alice expresses her true attitude, Bob will express whatever attitude Alice expressed, no matter his true attitude. 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. The unique equilibrium outcome subject to the intuitive criterion is for Bob to express Alice’s expressed attitude.

Notice that there are two equilibrium dynamics which depend on second-order uncertainty, a result that holds in the more general model. Without second-order uncertainty, no individuals reveal their true attitude – they all express whichever true attitude is more widespread in the population from which they know they are drawn.\(^9\) When second-order uncertainty is highest, 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.\(^10\) In the current example, this run is of size one. The result for the general model is presented in section 3.

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, or act reluctantly, and most individuals reach the end of the game believing most are not acting reluctantly. In the current example, this would require both individuals to be anti-segregationist and express a pro-segregationist attitude, or viceversa.

The probability of pluralistic ignorance is zero with the highest second-order uncertainty – since Alice always expresses her true attitude, in all realizations of true attitudes at least half the individuals do not act reluctantly. However, the probability of pluralistic ignorance is positive when there is no second-order uncertainty. When both individuals are anti-segregationist (an event with

\(^9\)The logic of the model is symmetric: I could have defined a lack of second-order uncertainty as a situation where individuals were sure that in the population most were anti-segregationist. We would then have found an equilibrium where all expressed an anti-segregationist attitude.

\(^10\)The uninformative context dynamic is analogous to herding (Banerjee, 1992) or information cascades (Bikhchandani, Hirshleifer and Welch, 1992). In the model, individuals may herd on what they think is the majority true attitude, which may lead a majority to express the minority true attitude. Unlike in a standard herding model, individuals are concerned with what their expressed attitude reveals to others about their true attitude, of which they are perfectly informed. Whereas group size does not affect the probability of herding in the standard model, it affects the probability of pluralistic ignorance.
probability), they will both act reluctantly by expressing pro-segregationist attitudes. Moreover, since neither individual revealed information, at the end of the game they will continue to believe the other individual is most likely pro-segregationist, and therefore didn’t act reluctantly.

I will generalize the current example by allowing group size to be a parameter of the model, and having individuals care about how they expect they will be judged on average by others in a group. In particular, I show that in sufficiently small groups, the probability of pluralistic ignorance is lowest when second-order uncertainty is high. This is the first half of the main result of the paper. The second half is that, for sufficiently large groups, the probability of pluralistic ignorance is lowest when second-order uncertainty is low. This result is presented in section 4.

Summary of and intuition behind the main result. The intuition of the main result can be stated simply. When second-order uncertainty is high, the first movers express their true attitude as they believe they are in the majority, and the rest herd. The first movers who express their true attitude are often more than half of a small group and do not act reluctantly. The herders are often more than half of a large group and sometimes do act reluctantly. When second-order uncertainty is low, everyone expresses the same attitude since they have strong priors over the majority true attitude in the population. The probability of pluralistic ignorance is the probability that that attitude differs from the majority attitude in the group. But by the law of large numbers, this probability converges to zero as the size of the group becomes arbitrarily large.

2 Setup

There is a group with \( I \) individuals and \( 2 + I \times 2 \) periods, with generic individuals \( i, j, k, l \in \{1, 2, ..., I\} \equiv I \), 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. Individuals are endowed with a privately observed type \( \theta_i \in \{H, L\} \). Types will serve as a more general term for the ‘true attitudes’ I have discussed so far. The state of the world \((\theta_1, \ldots, \theta_I)\) is a realization of all individuals’ types. It is natural to define the state of the world in this way since individuals will
care about the majority type in their group.

[Figure I about here]

Figure I shows a graphical representation of how the state of the world is assigned. Nature selects a population $\psi \in \{H, L\}$ from which individuals are drawn. Nature selects the population $\psi$ with probability $\chi \equiv P(\psi = H) \in [0, 1]$. Probability $\chi$ captures second-order uncertainty, or the commonly held priors over the population from which the group is drawn. I will refer to $\chi$ as the **context**. Individuals are randomly drawn from the selected population, and are randomly indexed from 1 to $I$. Conditional on the population $\psi$ being equal to $\theta \in \{H, L\}$, an individual’s type is equal to $\theta$ 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 $\theta$. Since $\pi$ 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 $\psi$ is not observed and an individual’s type $\theta_i$ is private information, both the context $\chi$ and the precision $\pi$ are commonly known. Notice that types are informative about the population.

Once the group is assigned, individuals take turns choosing actions and judging each others’ actions. In period $i$, individual $i$ chooses $a_i \in \{H, L\}$ after observing the history of play $h_i = \{a_1, a_2, ..., a_{i-1}\}$, with $\mathcal{H}$ the set of possible histories. I will refer to $a_i$ as an **action**, a more general term than ‘expressed attitude’, which I have used so far. Figure II denotes the timeline.

For compactness, I will say individual $i$ **chooses his type** if $a_i = \theta_i$. Further, if $x \in \{H, L\}$, **not-**$x$ is the value other than $x$ in the set, or $\{H, L\} \backslash x$.

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. These stylized dynamics can be used to study both public opinion formation and conversations.

[Figure II about here]
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 $J_{j,i}(a_i) \in \{0, 1\}$ 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:

$$a^*_i = \arg \max_{a_i \in \{H, L\}} \mathbb{E}u(a_i; \theta_i, h_i, I) =$$

$$\arg \max_{a_i \in \{H, L\}} \mathbb{I}\{\theta_i = a_i\} + \frac{\beta}{I - 1} \sum_{j \neq i} \mathbb{E}\left(J_{j,i}(a_i) \mid h_i, \theta_i\right)$$

(1)

with $\mathbb{I}\{\cdot\}$ the indicator function. Since we will focus on pure strategies, indifference is a knife-edge scenario which I’ll ignore.\(^{11}\) I refer to an individuals’ expectation over average judgments as his social expectation, and to $\beta$ as the weight on social expectations. I assume $\beta > 1$ – individuals would always choose their type if $\beta \in [0, 1]$. Concerns over social expectations can be justified as a reduced-form reputational concern.\(^{12}\) Note that I will generally use $i$ for the individual acting in the current period, or the speaker, and will use $j$ for a judge of $i$. To avoid confusion, I will refer to speakers in the masculine gender and judges in the feminine gender.

When judge $j$ judges speaker $i$, judge $j$ must decide whether $i$’s type matches her own type ($J_{j,i} = 1$) or not ($J_{j,i} = 0$). Judge $j$ sets $J_{j,i} = 1$ if it is most likely that $i$’s type matches her own, or $P(\theta_i = x \mid h_i, a_i, \theta_j = x) > 1/2$. If $j$ believes it is more likely that their types do not match, or $P(\theta_i = x \mid h_i, a_i, \theta_j = x) < 1/2$, then $J_{j,i} = 0$. As a shorthand, whenever there is no ambiguity I say $j$ believes types match if $J_{j,i} = 1$, and $j$ believes types do not match if $J_{j,i} = 0$. When there are no mixed strategies in equilibrium, the intermediate case of $P(\theta_i = x \mid h_i, a_i, \theta_j = x) = 1/2$

\(^{11}\)In the simple herding or information cascades model, indifference is common with pure strategies. Individuals only care about choosing a binary action that matches a binary and equiprobable state of the world, and everyone receives a binary signal that is equally informative. Therefore, observing an equal amount of signals leads to indifference, and specifying a tie-breaking rule is important – see, e.g., Banerjee (1992). In my case, indifference is uncommon because individuals are biased towards choosing their type, and because the information structure is asymmetric for all $\chi \neq 0.5$.

\(^{12}\)A concern for average judgments can be justified by assuming that speakers want to be judged positively by most individuals in the group, or equivalently, by as many as possible. Alternatively, suppose that a random subset of the judges decide whether to develop a relationship with the speaker, which is captured by judging the speaker positively. Then the concern over average judgment can be justified as a concern over the expected value of future interactions.
is a knife edge scenario.¹³ Judgments are not observed by others, perhaps because judges do not want to or cannot make them immediately public. Note that as long as the judgment is made after a speaker has acted, its exact timing does not matter. A judge’s judgment compares two types: her own, which she knows perfectly, and the speaker’s. To form beliefs about the speaker’s type, the judge uses the speaker’s action considering what he knew when making a decision.

Types impact utility through two channels. Since they are signals about the population, they affect judges’ judgments and speakers’ social expectations. In addition, they directly affect speakers’ utility through the first summand in (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.¹⁴

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, speaker i would not face any uncertainty in 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.

¹³A natural way to motivate this judgment function is that a positive judgment is a decision to engage with the speaker. For example, it may be a decision to establish a future relationship with the speaker, or simply to like or pay attention to him. The judge wants to engage with the speaker if and only if they are of the same type. Then she maximizes her expected utility:

$$\max_{J_j, i \in \{0, 1\}} P(\theta_i = x | h_i, a_i, \theta_j = x)J_{j,i} + P(\theta_i \neq x | h_i, a_i, \theta_j = x)(1 - J_{j,i})$$

In a mixed strategy Nash equilibrium, some judges would be indifferent in their judgment and would randomize in the natural way.

In Bursztyn, Egorov and Fiorin (2017), where individuals also care about how they are judged by others, they care directly about $P(\theta_i = x | h_i, a_i, \theta_j = x)$. Since $J_{j,i}$ is a discrete approximation, I conjecture that this modification would lead to qualitatively similar predictions, but would complicate the calculations significantly.

¹⁴Ross, Greene and House (1977) coined the term ‘false consensus effect’ to refer to the finding that individuals overestimate the proportion of others’ attitudes that matches their own expressed attitude. Mullen et al. (1985) provide a meta-analysis of 115 tests of this effect, finding robust consistent evidence. As Dawes (1989) pointed out, the effect may be driven by rational updating – using one’s own preference to infer what others prefer. This argument is echoed by our setup. More recent experimental work has studied the rationality of this effect (Engelmann and Strobel, 2000, 2012), and theoretical work has considered its impact on social learning (Gagnon-Bartsch, 2017).
2.1 Strategies and equilibrium

Denote by $G \equiv \{\chi, \pi, \beta, I\}$ the game defined above with context $\chi$, precision $\pi$, weight on social expectations $\beta$ and group size $I$. A pure strategy of $i$ is $\alpha_i : H \times \{H, L\} \rightarrow \{H, L\}$, which maps a history $h_i$ and a type $\theta_i$ to an action. I will be interested in Perfect Bayesian Equilibria, and use the intuitive criterion to refine out-of-equilibrium beliefs (Fudenberg and Tirole, 1991). The intuitive criterion posits that, if an action is chosen that no type would choose on the equilibrium path of play, individuals will infer that the deviation came from a type who would deviate for some out-of-equilibrium beliefs. I will refer to a Perfect Bayesian equilibrium with the intuitive criterion refinement as a PBE, and a strategy chosen in a PBE as a PBE strategy.

In Appendices A and B I provide results that are analogous to those presented in the body of the text but without the further equilibrium restrictions I am about to impose. The Appendix provides the full argument as to why these restrictions are reasonable. Here I will give a short justification.

A PBE strategy for speaker $i$ is **non-reverse** if the probability that speaker $i$ of type $\theta \in \{H, L\}$ chooses action $\theta$ in equilibrium is greater or equal to the probability that speaker $i$ of type not-$\theta$ chooses the same action $\theta$. Otherwise, we say the PBE strategy is a **reverse** strategy. Reverse strategies are unintuitive, do not always exist, and when they do there is often a non-reverse version of the strategy – for example, there is always a separating strategy if there is a reverse separating strategy where type $\theta_i$ chooses action not-$\theta_i$. A possible exception is ‘doubly mixed strategies’, where both types of a speaker randomize – there are no non-reverse doubly mixed strategies, but I cannot rule out reverse doubly mixed strategies. For simplicity and succinctness, we rule out reverse strategies. If a PBE is made up of non-reverse PBE strategies, we say the PBE is non-reverse.

I say that $k$ **pools on** $x \in \{H, L\}$ at history $h_k$ if he chooses $x$ no matter his type: $\alpha_k(h_k, \theta_k) = x$ for all $\theta_k$. When it does not lead to confusion, I will simply say that $k$ pooled. Notice that reverse PBE strategies do not rule out any pooling strategies. 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 say speaker $i$ **separates** at history $h_i$ if $i$ takes the action that matches
his type: \( \alpha_i(h_i, \theta) = \theta \) for \( \theta \in \{H, L\} \). I call \( k \) a **revealer** at history \( h_i \) if \( k \) has already made a decision \((k < i)\) and separated. **Semi-pooling on** \( x \) is defined in the obvious way for a non-reverse strategy. If it does not lead to ambiguity, I will not mention the history when referring to revealers and withholders, or to separating and pooling strategies.

Given that we are interested in equilibrium misaggregation of information, we make a conservative assumption if we restrict ourselves to selecting equilibria where individuals’ actions reveal the most information about their type. Given a set of PBE strategies for speaker \( i \), the **informative** PBE strategies are those which reveal most information about \( i \). That is, each action that \( i \) takes on the equilibrium path of play with an informative PBE strategy updates judges’ priors more than any other PBE strategy: \( |P(\theta_i = 1 \mid h_i, a_i) - P(\theta_i = 1 \mid h_i)| \) is largest with an informative PBE strategy for all \( a_i \in \{H, L\} \) on the equilibrium path of play. I will call an non-reverse PBE **informative** if the PBE strategies are informative. If separating is a PBE strategy, then it is the only informative PBE strategy. If semi-pooling is a PBE strategy and separating is not, pooling is not an informative PBE strategy unless we restrict to pure strategies.

**Definition 1.** An equilibrium is an informative, non-reverse Perfect Bayesian Equilibrium satisfying the intuitive criterion for refining out-of-equilibrium beliefs, which selects pure strategies whenever they exist. A strategy chosen in equilibrium is an **equilibrium strategy**.

### 3 Equilibrium Dynamics

This section presents the first main result of the model. After laying some groundwork, I describe the equilibrium dynamics of the model. This result is the cornerstone for the rest of the results. The proof is in Appendix A. In section 4, I give a formal definition of pluralistic ignorance and show how the probability of pluralistic ignorance depends on group size and context.

I first define informative and uninformative contexts, which are subsets of \( \chi \) I will focus on. For the remainder of the paper I focus on the range \( \chi \geq 1/2 \). Given the symmetry of the setup, this is without loss of generality.
3.1 Informative, Semi-Informative and Uninformative Contexts

The context $\chi$, which again is $P(\psi = H)$, captures the common knowledge over the distribution from which the group’s types were drawn. Notice there are two levels of uncertainty: the population from which the group is drawn, and the individuals drawn from that population. When beliefs over the population $\psi \in \{H, L\}$ are arbitrarily strong, it means that individuals hold arbitrarily strong beliefs that other individuals in the group are drawn from $\psi$. I’ll focus on contexts where individuals are very certain or very uncertain of the majority type.

**Definition 2.** In game $G$ with $\beta > 1$, the context $\chi$ is **informative** if and only if individual $i$ of type $L$ strongly believes most judges are of type $H$:

$$P(\theta_j = H \mid \theta_i = L) > \frac{\beta + 1}{2\beta} \in (1/2, 1)$$

The context $\chi$ is **semi-informative** if and only if, at the beginning of the game, individual $i$ of type $L$ weakly believes most judges are of type $H$:

$$\frac{\beta + 1}{2\beta} > P(\theta_j = H \mid \theta_i = L) > 1/2$$

The context $\chi$ is **uninformative** if and only if, at the beginning of the game, individual $i$ of type $x$ believes most judges are of type $x$, for either $x \in \{H, L\}$:

$$P(\theta_j = H \mid \theta_i = L) < \frac{1}{2} \leq P(\theta_j = H) < P(\theta_j = H \mid \theta_i = H)$$

A context is not informative, or is a **non-informative** context if it is either uninformative or semi-informative.

By the symmetry of the game, the definition of an informative context is without loss of generality – a context where speaker $i$ of type $H$ would believe most judges are of type $L$ is defined analogously, and the results below would apply symmetrically. The same is true for semi-informative contexts.
Unininformative and informative contexts are of particular interest, since they respectively capture high uncertainty and low uncertainty about the distribution of types from which the group was drawn. An informative context captures a perceived certainty over the distribution of types. This may be due to several reasons: attitudes may have been stable over time; there may be cultural homogeneity\(^{15}\); or there may be much public information about the distribution of attitudes (such as through surveys for large groups or because of familiarity in small groups).\(^{16}\) For example, the Jim Crow era in the South of the U.S. was an informative context for whites regarding segregation, since there was a strongly perceived pro-segregation consensus. An uninformative context would arise from converse reasons. The Civil Rights movement and its aftermath were a period of flux regarding segregationist attitudes in the U.S., capturing an uninformative context.

As I show in Appendix A, PBE strategies are pure in informative and uninformative contexts. The following provides conditions for informative contexts to exist, and characterizes the uninformative and semi-informative contexts, which always exist:

**Lemma 1.** Consider a game \(G\). The context \(\chi\) is uninformative if and only if \(\chi \in [1/2, \pi]\).

If the precision \(\pi > (1 + \beta)/2\beta\), then the context \(\chi\) is informative as long as it is greater or equal to:

\[
\chi^i(G) \equiv \frac{\pi(\beta(2\pi - 1) + 1)}{(1 - \pi)(\beta(2\pi - 1) - 1) + \pi(\beta(2\pi - 1) + 1)} \in (\pi, 1)
\]

A context \(\chi\) is semi-informative if and only if \(\chi \in (\pi, \chi^i(G))\).

If the precision \(\pi \leq (1 + \beta)/2\beta\), then there are no informative contexts, and a context is semi-informative if and only if \(\chi > \pi\).

When it does not lead to ambiguity, I will omit the argument of \(\chi^i\).

---

\(^{15}\)Note that cultural heterogeneity may mean two things in the model. It may be that it is well known that the population is split in its types, which is captured by \(\pi\) close to 0.5. This is not the cultural heterogeneity to which I’m referring. The second interpretation is that, due to differences in some known characteristics, it is hard to make strong predictions about the distribution of types on the relevant characteristic. This is captured by \(\chi\) close to 0.5.

\(^{16}\)Another source of a strong belief in the distribution of types would be social entrepreneurs who disseminated information about the distribution of types. However, since social entrepreneurs may try to disseminate misinformation, I consider this channel separately in section 7.
3.2 Characterizing Equilibrium Dynamics

In order to keep track of the information that has been revealed by individuals’ actions, I follow Ali and Kartik (2012) and define 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 type $L$.

I now introduce an assumption that will allow me to focus on pure strategies even in semi-informative contexts.

**Assumption 1.** If at history $h_i$ of a game $G$ with a semi-informative context, (a) speaker $i$ of type $L$ expects others to be most likely of type $L$, or $\mathbb{E}_{j \neq i}(P(\theta_j = L \mid h_i, \theta_i = L)) > 1/2$; (b) a judge $j$ without private information believes the speaker $i$ is most likely of type $H$, or $P(\theta_i = H \mid h_i) > 1/2$; and (c) $-\Delta(h_i) > 0$, then $-\Delta(h_i)/(I-1) < 1/\beta$.

Whenever possible, I will make Assumption 1. In Appendix A I show that conditions (a) and (b) imply condition (c) and a semi-informative context, that (a) and (b) are the only conditions that can lead to a unique semi-separating equilibrium, and that if $-\Delta(h_i)/(I-1) < 1/\beta$ holds, the unique equilibrium strategy is to separate.

**Result 1.** Consider a game $G \equiv \{\chi, \pi, \beta, I\}$ and suppose $\pi > (\beta + 1)/2\beta$. Then all speakers pool on $H$ if $\chi > \chi_i(G)$. If $\chi \leq \chi_i(G)$, separating is a unique equilibrium strategy for speaker $i$ unless, given a history where all players have separated, the lead for action $a$ is greater or equal to $n(a, h_i; G) \geq 1$. In that case, pooling on $a$ is an equilibrium strategy for speakers $k \geq i$, and separating is not an equilibrium strategy. For $\theta \in \{H, L\}$, the threshold $n(\theta, h_i; G)$ decreases in $\beta$ and in $\pi$. For a given $\Delta(h_i)$, $n(\theta, h_i; G)$ weakly increases in $i$. Further, $n(H, i; G)$ and $n(L, i; G)$ decrease and increase, respectively, in $\chi$, with $n(H, i; G) = n(L, i; G)$ for $\chi = 1/2$.

Lemma 2 in Appendix A provides a stronger version of Result 1, and uses a weaker solution
The equilibrium dynamics of an informative context differs from those of a non-informative context. When the context is informative, the first speaker will strongly believe most judges are of type $H$, independent of his type. His belief is strong enough that he will pool on $H$ out of social expectations concerns. Since speaker 2 does not learn anything from speaker 1’s action, he will have the same incentives as speaker 1 and will therefore also pool on $H$. So will all other speakers.

When the context is non-informative, individuals at the beginning of the game rely heavily on their types to form expectations about the group. Compare the beliefs of an individual of type $\theta$ to an individual of type not-$\theta$. Type $\theta$ 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. In a non-informative context, this difference makes speaker 1 willing to reveal his type by choosing his type – he believes that many judges will judge him positively for doing so. Speaker 1 therefore separates, which gives speaker 2 higher incentives to pool on speaker 1’s type. Once enough speakers create a sufficiently large lead for some action, all speakers pool on that action. The thresholds for pooling are lower when the weight on social expectations $\beta$ is higher, and when the true population is very informative about the distribution of types in the group (i.e., $\pi$ is higher). The last result follows since there are fewer individuals whose type is unknown, and therefore whose type can be updated using information about the population. For a given action lead $\Delta(h_i)$, the thresholds for pooling are weakly higher the more individuals have taken an action.

Result 1 is an incomplete description of equilibrium dynamics, since for some histories in a semi-informative context, the unique equilibrium strategy is a mixed strategy – that is, Assumption 1 cannot hold. Nevertheless, for an analysis restricted to histories with pure strategies, semi-informative contexts are intermediate between uninformative and informative contexts. As the context increases within the range of non-informative contexts, it requires a smaller action lead of $H$ for the rest of the speakers to pool on $H$, and a larger action lead of $L$ for the rest of speakers to pool on $L$.

Although I will not consider histories of play in semi-informative contexts where mixed strate-
gies are the unique equilibrium strategy, I conjecture that the dynamics would be similar. A mixed strategy arises when an individual of type $\theta$ has observed enough signals of type not-$\theta$ to deviate from a separating strategy and from a strategy of pooling on not-$\theta$. The equilibrium strategy would be to semi-pool on not-$\theta$. With a higher probability than in a separating strategy, the speaker chooses not-$\theta$. By choosing not-$\theta$, judges believe he is more likely of type not-$\theta$ than before his action, although beliefs are updated less than if he were separating and chose not-$\theta$.

The uninformative context gives a lot of influence to the first movers. A social planner interested in getting most speakers to choose the same action may want to exploit the first movers’ influence. In section 7.2 I consider what would happen if the social planner co-opted first movers to choose her preferred action.

The speaker’s concern over expected judgments, together with the fact that judgments are silent, imply that in equilibrium a speaker will choose the same action whether he acts alone or at the same time as other speakers. A simultaneous-move version of game would then result in all speakers following whichever strategy is taken by the first mover of the sequential-move version. I further discuss the difference between the two versions in Supplementary Materials 1, where I consider the extended model of section 5.

4 Pluralistic Ignorance in Large and Small Groups

Result 1 described equilibrium dynamics for informative and non-informative contexts. In this section, I relate these dynamics to pluralistic ignorance. I will first define pluralistic ignorance as a situation where most individuals are reluctant choosing the action they mistakenly believe is the majority type. In subsection 4.2 I will go through a few specific cases – that of groups of size 2, 3 and $\infty$ – to illustrate the impact of group size on dynamics and pluralistic ignorance. In subsection 4.3 I show the main result of the paper: that the probability of pluralistic ignorance arising in a group depends on the context and the group size. The proof is in Appendix B.
4.1 Defining Pluralistic Ignorance

I say speaker $i$ of type $\theta_i$ acts reluctantly at history $h_i$ if his equilibrium action differs from his type: $\theta_i \neq \alpha_i^*(\theta_i, h_i)$. Speaker $i$ would not act reluctantly if he did not care sufficiently about social expectations ($\beta \in [0, 1]$). To illustrate reluctance, take the working example of segregation: whites who act reluctantly express support for segregation when they do not support it, or express that they do not support segregation when they do support it.

Definition 3. There is pluralistic ignorance for realization $\overline{\theta} \equiv (\theta_1, \theta_2, ... \theta_I)$ in an equilibrium of game $G$ if and only if (a) most agents act reluctantly: $P(\theta_i \neq \alpha_i^*(\theta_i, h_i) \mid \overline{\theta}) > 1/2$, and (b) at the end of the game, most agents believe most others did not act reluctantly:

$$E_i \left( \sum_{j \neq i} \frac{P(\theta_j \neq \alpha_j^*(\theta_j, h_j) \mid \theta_i, h_I, a_I)}{I-1} \bigg| \overline{\theta} \right) < 1/2.$$

If there is pluralistic ignorance, most individuals in the group act reluctantly, but believe most others are not acting reluctantly. Pluralistic ignorance among whites would be either for most to reluctantly support segregation and for most to believe that most supported segregation non-reluctantly, or for most to reluctantly oppose segregation and for most to believe that most opposed segregation non-reluctantly. Substituting the strict inequalities for weak inequalities in the definition leads to qualitatively similar results. The existence of pluralistic ignorance in the model is a corollary of Result 1.

Although the formal definition of pluralistic ignorance is novel, in section 9 I argue that it captures important features of a widespread, indeed the original, use of the term. Furthermore, I will compare the definition to past formalizations, and argue that those definitions do not capture the same features, or are complementary to my definition.

Since pluralistic ignorance is defined for a realization $\overline{\theta}$ of types, the probability of pluralistic ignorance for a game $G$ is the probability that types are realized in a way that results in pluralistic ignorance.
4.2 Building Intuition with Groups of Size 2, 3 and ∞

In this section I will consider equilibrium dynamics in groups of size 2, 3, and for arbitrarily large groups. These three cases are sufficient for illustrating the forces behind the main results. In particular, I’ll show how equilibrium dynamics affect the probability of pluralistic ignorance differently in small and large groups. The three cases also provide counter-examples of what may be intuitive predictions of what the model would predict. I’ll show that the probability of pluralistic ignorance is not monotonic in the precision π or the context χ. In fact, the probability of pluralistic ignorance may be lowest for some intermediate value of χ between 0.5 and 1. Further, I’ll show that increasing the group size may increase the number of individuals who reveal their type.

For all I, the first speaker separates in the uninformative and semi-informative contexts, and pools in the informative contexts. Therefore, I will only consider equilibrium strategies from the second speaker on.

- I = 2. Suppose the context is uninformative or semi-informative. The second speaker then knows the type of the first speaker, who is his only judge. In an uninformative context, the second speaker will pool on the first speaker’s type. In a semi-informative context, the second speaker pools on the first speaker’s type if the first speaker is of type H. If the first speaker is of type L, the second speaker semi-pools on L.17 This example shows that semi-pooling may be the unique equilibrium strategy for any β.

Regardless of the equilibrium strategy of the second speaker, the probability of pluralistic ignorance is zero – the first speaker does not act reluctantly, and he makes up half of the group. However, the probability of pluralistic ignorance is positive in an informative context. When both speakers choose action H independent of their type, there is a probability χ(1 − π)^2 + (1 − χ)π^2 that both types are acting reluctantly. This probability declines as χ increases.

---

17This follows since (a) the second speaker knows the first speaker is of type L, (b) when the first individual judges the second, her prior is that the second individual is most likely of type H, independently of the first individual’s type, and (c) −Δ(h2)/(I − 1) = 1 > 1/β. Therefore, Assumption 1 cannot hold. Conditions (a) and (b) imply that pooling is not an equilibrium strategy, since no judge would believe types matched if the speaker pooled. Condition (c) implies that speaker of type H would deviate from a separating strategy.
Notice that the probability of pluralistic ignorance is non-monotonic in $\chi$: it is equal to 0 for uninformative and semi-informative contexts, reaches a maximum level for the minimum value of $\chi$ such that the context is informative, and is positive but decreasing for larger values of $\chi$. Similarly, we can use Lemma 1 to show the probability of pluralistic ignorance is non-monotonic in $\pi$. For $\pi = 0.5$, the probability of pluralistic ignorance is zero since the first speaker separates – for any $\chi$, he believes the second individual is equally likely to be of either type. By Lemma 1, if $\pi \in ((1 + \beta)/2\beta, \chi)$ the context is informative, so the probability of pluralistic ignorance is positive. If $\pi > \chi$, the context is uninformative or semi-informative, so the probability of pluralistic ignorance is zero.

- $I = 3$. Consider first uninformative and semi-informative contexts. For a low enough weight on social expectations $\beta$, the second speaker would always separate, and the probability of pluralistic ignorance would be 0 as in the case of $I = 2$. However, the probability of pluralistic ignorance is positive for high values of $\beta$ – the second and third speaker would pool on $H$ if the first speaker was of type $H$. Increasing the group size from 2 to 3 then weakly increases the probability of pluralistic ignorance. Moreover, the probability of pluralistic ignorance may increase in the context $\chi$ within the range of uninformative contexts. Suppose that for some $\chi'$ in the range of uninformative contexts, the second and third speaker pool on either type of the first speaker. The condition for this to hold can be derived by using Bayes’ rule and algebraic manipulations:

$$\frac{2 - \beta}{\beta} < 2(\pi - 1)(2\chi' - 1) < \frac{\beta - 2}{2} \quad (2)$$

The probability of pluralistic ignorance would then be $\pi(1 - \pi)^2 + (1 - \pi)\pi^2$. Further suppose that, consistent with Result 1, an uninformative context $\chi'' > \chi'$ leads second and third speakers to pool on $H$ if the first speaker is of type $H$, but second speakers to separate
if the first speaker is of type $L$:

$$\frac{\beta - 2}{\beta} < (2\pi - 1)(2\chi'' - 1) \quad (3)$$

The probability of pluralistic ignorance would then be $\chi''(\pi(1 - \pi)^2) + (1 - \chi'')(\pi)^2$, which is smaller than $\pi(1 - \pi)^2 + (1 - \pi)^2$ for all $\chi''$. In fact, the probability of pluralistic ignorance may be smaller than in an informative context. As in the case of $I = 2$, the probability of pluralistic ignorance in an informative context is lowest when $\chi = 1$, with the probability equal to $(1 - \pi)^3 + 3\pi(1 - \pi)^2$. This probability may be higher than the probability of pluralistic ignorance with context $\chi''$. An example of parameter values that fulfill this condition along with (2), (3) and $\chi'' < \pi$ (so the context is uninformative) are $\beta = 2.5$, $\chi' = 0.6$, $\chi'' = 0.7$ and $\pi = 0.8$. Similar conclusions would follow if $\chi''$ was in a semi-informative context, although the probability of pluralistic ignorance will depend on the probability with which the second speaker of type $L$ would randomize in a semi-pooling strategy. We can then conclude that the minimum value of the probability of pluralistic ignorance need not be reached in the extreme values of $\chi = 0.5$ or $\chi = 1$.

- $I \to \infty$. Begin with uninformative and semi-informative contexts. There are no semi-pooling strategies in equilibrium since, for any history $h_i$, $\lim_{I \to \infty} \Delta(h_i)/(I - 1) = 0 < 1/\beta$. Lemma 6 and its proof, presented in Appendix B, show that if $\pi \leq (\beta + 1)/2\beta$, then speakers separate for any $\chi$ when $I$ is arbitrarily large, and therefore there is no pluralistic ignorance.\(^{18}\) When $\pi > (\beta + 1)/2\beta$, however, speakers will pool after some histories of play. Since $I$ is arbitrarily large, this implies that there is a positive probability of pluralistic ignorance.

As before, the probability of pluralistic ignorance decreases with $\chi$ when $I \to \infty$ and the context is informative. Recall that in these contexts all individuals are choosing $H$. By the law of large numbers, the majority type of the group is equal to the majority type of the

\(^{18}\)The proof of Lemma 6 shows that the threshold for pooling on $L$ becomes arbitrarily large as $I$ goes to infinity, and a similar logic applies to the threshold for pooling on $H$. 
population for an arbitrarily large $I$. Therefore, for a context $\chi'$ in the range of informative contexts, the probability of pluralistic ignorance is simply $1 - \chi'$. But then the probability of pluralistic ignorance can be made arbitrarily small for a sufficiently high $\chi'$. In particular, there is some $\hat{\chi}$ in the range of informative contexts such that the probability of pluralistic ignorance is lower for any $\chi \in [\hat{\chi}, 1]$ than for any other context.

### 4.3 Context, Group Size and Pluralistic Ignorance

In this section I will present the main result, which sheds light on how group size and context affect pluralistic ignorance. I will assume that $\pi > (\beta + 1)/2\beta$, which implies that an informative context exists.

**Result 2.** Suppose $\pi > (\beta + 1)/2\beta$. Then there are thresholds $I(G), I(G), \tilde{\chi}(I)$ and $\tilde{\tilde{\chi}}(I)$, with $\tilde{I}(G) > I(G)$, such that:

- When $I \geq \tilde{I}(G)$, the probability of pluralistic ignorance is smaller for any $\chi \geq \tilde{\chi}(I) \in [\chi^i, 1]$ than for any other context.

- When $I \leq I(G)$, the probability of pluralistic ignorance is 0 only for contexts $\chi \leq \tilde{\tilde{\chi}}(I) \in [0.5, \chi^i]$.

- $I(G)$ and $\tilde{I}(G)$ weakly increase as $\beta$ decreases.

When the group is small and the non-informative context is sufficiently close to 0.5, the first movers who reveal their type may be more than half of the group, in which case there is no pluralistic ignorance. However, the probability of pluralistic ignorance is positive with an informative context since everyone chooses type $H$ and believes most are of type $H$, even if most are of type $L$. When the group is large, if everyone chooses action $H$, then the probability of pluralistic ignorance tends from above to $1 - \chi$. With a high enough $\chi$ and a large enough group, the probability of pluralistic ignorance can then be made arbitrarily small. If the group is large and the context is non-informative, there is a positive probability of pluralistic ignorance: speakers may herd on the
majority type of first movers which happens to be the minority type of the group. Therefore, the smallest probability of pluralistic ignorance in a large group comes from an informative context sufficiently close to 1.

Group size is exogenously determined in the model. It depends on how many individuals can observe and judge an action – as discussed in the introduction, a Twitter post can be seen by a much larger group than a comment in a private conversation. The model makes the convenient simplification that the set of speakers equals the set of judges. This simplification is non-essential, however. Similar results would hold if in small groups there were few speakers and judges, but the set of the judges grew faster than the set of speakers as the group size increased. This modification provides a better fit for thinking about public opinion expressions, where not everyone makes a public statement.

In Lemma 7 of Appendix B I present a more thorough description of how the probability of pluralistic ignorance is affected by the parameters in the model. I also discuss limitations to further extending the model, and provide further numerical examples.

5 Pluralistic Ignorance and Collective Action

In this section I modify the utility function to allow for externalities. This will allow me to study the relationship between pluralistic ignorance and collective action. Individuals’ actions will now affects others’ payoffs directly. The modified utility function is

\[
\max_{a_i \in \{H,L\}} 1 \{\theta_i = a_i\} + \gamma_{t(i)} \sum_{k \neq i} 1 \{\theta_i = a_k\} + \frac{\beta}{I-1} \sum_{j \neq i} \mathbb{E} \left( J_{j,i} | h_i, \theta_i \right)
\]

(4)

The first two summands are the individuals’ material payoff. The second summand is new, equal to the proportion of others’ actions that matches the individual’s type. A natural interpretation for this is that the action provides a public good to the individual, with the direction of the public good equal to his type. So if the action is to protest in opposition of segregationist housing, those who oppose segregationist housing want more people to protest, while those who support it
want less people protesting. I take the average action in the second summand instead of the sum so that the impact of others’ actions does not rise mechanically with group size.

Notice that the weight on others’ actions, $\gamma_{\theta(i)} > 0$, depends on the individual’s type. This will allow us to consider asymmetries in the externalities of the actions: $\gamma_H > \gamma_L$ indicates that type $H$ puts relatively more weight on others’ actions matching his type than does type $L$. For example, it may be that anti-segregationists benefit more from opposition to segregation than pro-segregationists do from support for segregation, or that they take into account the welfare of those who are being segregated. We then define a game $\hat{G} \equiv \{\chi, \pi, \gamma_L, \gamma_H, \beta, I\}$ where the utility function is given by (4), and the rest of the setup is the same as before. The analogous to Assumption 1 holds, which replaces game $G$ for game $\hat{G}$. Notice that if $\gamma_H = \gamma_L = 0$, we are back to the original setup from section 2.

With this new setup, we can establish the following result.

**Result 3.** Consider a game $\hat{G} \equiv \{\chi, \pi, \gamma_L, \gamma_H, \beta, I\}$. Then there is a $\hat{\pi} \in ((\beta + 1)/2\beta, 1)$ such that for all $\pi > \hat{\pi}$, all speakers pool on $H$ if $\chi > \chi^i(\hat{G})$ for some $\chi^i(\hat{G}) \in (0.5, 1]$. If instead $\chi \leq \chi^i(\hat{G})$, separating is a unique equilibrium for speaker $i$ unless, given a history where all players have separated, the lead for action $a$ is greater or equal to $\hat{n}(a, h_i; \hat{G})$. In that case, pooling on $a$ is an equilibrium strategy for speakers $k \geq i$, and separating is not an equilibrium strategy. For $\theta \in \{H, L\}$, the threshold $\hat{n}(\theta, h_i; \hat{G})$ increases in $\gamma_\theta$, and decreases in $\beta$ and in $\pi$. For a given $\Delta(h_i)$, $\hat{n}(\theta, h_i; \hat{G})$ weakly increases in $i$. Further, $\hat{n}(H, h_i; \hat{G})$ and $\hat{n}(L, h_i; \hat{G})$ decrease and increase, respectively, in $\chi$, with $\hat{n}(H, h_i; \hat{G}) = \hat{n}(L, h_i; \hat{G})$ for $\chi = 1/2$.

Consider game $G \equiv \{\chi, \pi, \beta, I\}$, where the parameters $\chi$, $\pi$, $\beta$ and $I$ are the same as in game $\hat{G}$. Then the threshold for pooling on $\theta \in \{H, L\}$ is larger in game $\hat{G}$ than in game $G$: $\hat{n}(\theta, h_i; \hat{G}) \geq n(\theta, h_i; G)$. The minimum informative context is weakly greater in $\hat{G}$ than in $G$: $\chi^i(\hat{G}) \geq \chi^i(G)$.

The proof uses a form of backward induction, and is in Supplementary Materials 3. For a given set of beliefs, the last speaker in game $\hat{G}$ with externalities faces the exact same incentives as the last speaker in an equivalent game $G$ without externalities. The next-to-last speaker then either
cannot influence the last speaker – in which case his incentives are the same as that of the next-to-
last speaker in game $G$ – or will raise the probability that the last speaker chooses $a$ if he chooses
$a$. But this is true of all speakers other than the last speaker, so all have weakly greater incentives
to choose the action that matches their type in game $\hat{G}$ than in game $G$.

We now turn to the question of how group size affects the probability of pluralistic ignorance
when individuals care about others’ actions, as in utility (4). Given the similarity between Results
1 and 3, as well as the fact that Result 1 is the basis for Result 2, we can derive the analogue of
Result 2 for when the utility function is (4).

**Result 4.** For a game $\hat{G}$ there is a $\hat{\pi}$ such that if $\pi \geq \hat{\pi} \geq (\beta + 1)/2\beta$, there are thresholds $I(\hat{G})$, $\chi(\hat{G})$ and $\chi(\hat{G})$, with $T(\hat{G}) > I(\hat{G})$, such that:

- When $I \geq T(\hat{G})$, the probability of pluralistic ignorance is smaller for any $\chi \geq \chi(\hat{G}) \in [\chi(\hat{G}), 1]$ than for any other context.
- When $I \leq I(\hat{G})$, the probability of pluralistic ignorance is 0 only for contexts $\chi \leq \chi(\hat{G}) \in [0.5, \chi(\hat{G})]$.
- $I(\hat{G})$ and $T(\hat{G})$ weakly increase as $\beta$ decreases.

The logic is basically the same as that of section 4, so I omit repeating the intuition or the proof.

### 5.1 Illustration of Groups where Actions Have Externalities

By allowing for externalities in the utility function, we can use the model to think of a broader set
of applications. Interpret the proportion of individuals who choose an action as the probability that
a public good is provided. Then the model can be used to think of large-group phenomena such
as collective action problems. For example, we may use the model to think of protests against a
regime, charity drives to find a cure for a disease, social movements to get a law passed, and so on.
It can also be used to think of small-group phenomena. A joint project – a business venture, going
on a date, going drinking with friends, starting a study group, a yearly gift exchange, organizing
a party – is more likely to begin within a group the more individuals express support for the idea. The model shows how these joint projects can be affected by misunderstandings.

We’ve just seen that the model can be extended to capture settings with collective action, and that pluralistic ignorance arises in similar conditions. Misunderstandings about others’ types may lead to actions that affect others’ utility directly. What are the social welfare implications of pluralistic ignorance, and how can we avoid inefficiencies that arise from pluralistic ignorance? To these questions we now turn.

6 Social Welfare Consequences of Pluralistic Ignorance

To understand whether pluralistic ignorance is generally good or bad, we must provide a definition of social welfare. A proper definition depends on whether social expectations have social or material consequences. I will define different social welfare functions based on different views of social expectations.

Suppose first that individuals care about how they think they are judged by others for some non-consequentialist reason. Under this interpration, others may not even be really judging them – it may all be in the individuals’ head. Alternatively, others may actually be judging them, but the judgment may be of no long-term consequence to the judges. If social expectations are non-consequential, then arguably social welfare should not take social expectations into account – for example, because what we really care about are the material consequences of actions. The social welfare function would then be $W_1(\theta_1, \ldots, \theta_I) \equiv$

$$\sum_{i \in I} \left[ \mathbb{1}\{\theta_i = a_i\} + \gamma_{\theta(i)} \frac{\sum_{k \neq i} \mathbb{1}\{\theta_i = a_k\}}{I - 1} \right]$$

Alternatively, we may care about including social expectations in the utility function, since we should arguably weight their hedonic impact. A social welfare function under this interpretation...
simply takes the sum of utilities, including all terms. That is, 
\[ W_2(\theta_1, \ldots, \theta_I) \equiv \sum_{i \in I} \left[ \mathbb{I}\{\theta_i = a_i\} + \frac{\gamma_{\theta(i)}}{I} \sum_{k \neq i} \mathbb{I}\{\theta_i = a_k\} \right] + \frac{\beta}{I} \sum_{j \neq i} \mathbb{E} \left( J_{j,i} \mid h_i, \theta_i \right) \]

A different way of thinking about social expectations is that they capture how individuals believe they will be judged by others when they take an action, and judgments have social or material consequences. For example, if \( k \) judges \( i \) negatively, \( k \) may not want to associate with \( i \) in a future interaction. Under this interpretation, actual judgments matter, not just how individuals expect they will be judged. The social welfare function would then be 
\[ W_3(\theta_1, \ldots, \theta_I) \equiv \sum_{i \in I} \left[ \mathbb{I}\{\theta_i = a_i\} + \frac{\gamma_{\theta(i)}}{I} \sum_{k \neq i} \mathbb{I}\{\theta_i = a_k\} \right] + \frac{\beta}{I} \sum_{j \neq i} J_{j,i} \]

Notice that in \( W_3 \), social welfare considers the actual judgments as opposed to the expected judgments considered in \( W_2 \).

We can now ask what the impact of pluralistic ignorance is on these different social welfare functions. Pluralistic ignorance is defined for a specific realization of types \( \bar{\theta} \) in a given game \( G \). In order to consider how variations in pluralistic ignorance impact social welfare, I will fix the realization \( \bar{\theta} \) and consider variations in the context \( \chi \).

**Result 5.** Consider a realization of types \( \bar{\theta} \equiv \{\theta_1, \ldots, \theta_I\} \) with majority type \( L \) in games \( \hat{G} \equiv \{\chi, \pi, \gamma_L, \beta, I\} \) and \( \hat{G}' \equiv \{\chi', \pi, \gamma_L, \gamma_H, \beta, I\} \).

Suppose equilibrium dynamics in game \( \hat{G} \) for realization \( \bar{\theta} \) are such that the first \( x \geq 0 \) individuals separate, the other \( (I - x) > I/2 \) pool on \( H \) and there is pluralistic ignorance. There are two alternative sets of assumptions regarding the equilibrium dynamics in game \( \hat{G}' \) for realization \( \bar{\theta} \). The first set of assumptions is that the first \( y > x \) individuals separate and the other \( I - y \) pool. The second set of assumptions is that the first \( y < x \) individuals separate and the other \( I - y \) pool on \( L \). With the second set of assumptions there is no pluralistic ignorance.

Then there exists a \( \Gamma_H(y, x, I, \bar{\theta}, \hat{W}, \beta, \gamma_L) \in (0, \infty) \) and a \( \hat{\beta}(y, x, I, \bar{\theta}, \hat{W}, \gamma_L, \gamma_H, I) > 1 \).
such that social welfare is weakly higher in game $\hat{G}$ if and only if (a) for any $\hat{W} \in \{W_1, W_2, W_3\}$, $\gamma_H - \gamma_L \geq \Gamma_H$, or (b) $\hat{W} = W_2$ and $\beta \geq \hat{\beta}$. The same result holds by swapping $H$ and $L$.

Result 5 considers a realization of types $\bar{\theta} \equiv \{\theta_1, \ldots, \theta_I\}$, and compares the equilibrium dynamics for this realization in two games $\hat{G}$ and $\hat{G}'$ that differ only in the context. In pluralistic ignorance most individuals are acting reluctantly, so by definition there are less individuals who choose the action that matches their type in a game $\hat{G}$ whose dynamics result in pluralistic ignorance than in a game $\hat{G}'$ whose dynamics do not. Therefore, the only way for pluralistic ignorance to result in a higher material payoff – again, the terms in the utility function other than social expectations – is for the externalities from choosing the minority type to be sufficiently greater than the externalities from choosing the majority type. By a similar argument, a dynamic that leads to pluralistic ignorance has a higher sum of material payoffs than a second dynamic that is identical except that the second has a smaller set of first movers who separate.

Pluralistic ignorance may also increase a social welfare of function if it includes individuals’ social expectations, as in $W_2$. This would happen if pluralistic ignorance leads individuals to believe they are judged more positively than they would otherwise (examples are provided in Supplementary Materials 4). Although a theoretical possibility, in practice it seems a hard sell to promote the welfare benefits of misinformation as a way to make people believe that they are liked more than they really are. Indeed, if what we care about is how people are actually judged by others, as in $W_3$, then this type of reasoning does not justify preferring an equilibrium dynamic that leads to pluralistic ignorance over one that does not.\(^{19}\) Analogous to before, a dynamic that leads to pluralistic ignorance has a higher sum of judgments than a second dynamic that is identical except that the second has a smaller set of first movers who separate.\(^{20}\)

If the externalities from the group’s majority type are greater or equal to those of the group

---

\(^{19}\)To see this, recall that in pluralistic ignorance the first $x \geq 0$ speakers reveal their type, and the other $I - x > I/2$ pool on the minority type – say, on $H$. Only type $L$ speakers among the first $x$ will be judged positively by most speakers – the rest will be judged negatively by most speakers. But if there is no pluralistic ignorance, then strictly more speakers will be judged positively by most judges – either because more reveal they are of type $L$ or because they pool on $L$ and others believe they are most likely of type $L$.

\(^{20}\)Since there are no other channels through which the equilibrium dynamics of game $\hat{G}$ can be better than the dynamics of game $\hat{G}'$, this establishes the Result.
minority type, pluralistic ignorance according to $W_3$ is triply inefficient – most individuals are acting reluctantly (first summand in $W_3$), their reluctant actions are minimizing externalities (second summand in $W_3$), and most individuals are judging most others negatively (third summand in $W_3$). An illustration of this type of triple inefficiency comes from the motivating example. Most whites in the U.S. in the sixties expressed that they did not favor segregation, that they thought most whites did favor segregation, and those who overestimated the amount of support for segregation were more supportive of segregationist housing policies. The interpretation yielded by the model is that whites were supporting a policy they themselves did not favor because they mistakenly thought most others favored the policy. Segregationist housing policies were then inefficiently supported, and whites negatively judged most other whites because they incorrectly believed that support for these policies came from an actual favorable attitude towards segregation.

I call the triple inefficiency of pluralistic ignorance a **perverse failure of collective action**. Individuals are deviating from the action that maximizes their material utility to act in a way that they think is cooperative, but in fact are minimizing social welfare. In standard examples of failures of collective action, individuals are tempted away from social-welfare-maximizing cooperative behavior by their selfish preferences (e.g. Olson, 1965, Hardin, 1982, Esteban, 2001, Siegel, 2009). Here, concerns over social expectations lead individuals away from their ‘selfish action’ (choosing their type), and into a worse type of inefficiency – one where most believe they are doing what others want.

### 7 Information Dissemination Policies

Suppose a social planner $P$ can intervene in the group dynamic by disseminating information about the group. What is the optimal information dissemination policy? To make this question more concrete, we must specify the social planner’s objective and her available information. In section 7.1 I consider a social planner who has private information about the population and must choose an announcement strategy. In section 7.2 I consider a policy of co-opting first movers to
take a specific action. In section 7.3 I consider a policy of varying the weight on social expectations to impact whether individuals choose their type.

### 7.1 Social Planner with Private Information About the Population

The social planner $\mathcal{P}$ will observe which population Nature selected from which to draw the group, but not the types of individuals themselves. The new timing of the game is captured in Figure III. Everything is the same as in section 5, except that there is a Period $-0.5$ after Nature selects a population and before Nature draws types from the selected population. In Period $-0.5$, the social planner $\mathcal{P}$ first privately observes the population Nature selected. With this information in hand, $\mathcal{P}$ decides whether to announce $A \in \{H, L\}$. I assume that $\mathcal{P}$ can perfectly commit \textit{ex ante} to an announcement strategy $A : \{H, L\} \rightarrow [0, 1]$, which is a function that goes from the population chosen by Nature to the probability that $\mathcal{P}$ announces $H$.

[Figure III about here]

Social planner $\mathcal{P}$ is interested in maximizing social welfare. I assume that $\mathcal{P}$’s objective function is either $W_1$ with $\gamma_H > 1$ and $\gamma_L > 1$, $W_2$ with $\beta$ low enough, or $W_3$. If I didn’t assume $\gamma_H > 1$ and $\gamma_L > 1$ for social welfare function $W_1$, it would always be socially optimal for speakers to choose their type. If I didn’t assume $\beta$ is small for social welfare function $W_2$, pluralistic ignorance may be socially optimal solely because individuals wrongly believe they are judged positively. This assumption thus rules out a weak social welfare argument for inducing pluralistic ignorance, as I argued in section 6.

Let $\hat{G}_\mathcal{P} = \{\chi, \pi, \gamma_L, \gamma_H, \beta, I\}$ be a game with a social planner $\mathcal{P}$. Take a realization of types $\theta$ with minority type $\theta \in \{H, L\}$. A pluralistic ignorance outcome where the majority choose $\theta$ is preferred by social planner $\mathcal{P}$ if the externalities of $\theta$ are large enough (Result 5). That is, there is some $\Gamma_\theta > 0$ such that for $\gamma_\theta - \gamma_{\text{not-}\theta} > \Gamma_\theta$, social planner $\mathcal{P}$ prefers equilibrium dynamics in which $\theta$ results in pluralistic ignorance to ones that do not. Therefore, as long as the difference in externalities $\gamma_\theta - \gamma_{\text{not-}\theta}$ are high enough, the social planner $\mathcal{P}$ would choose an announcement
strategy that maximizes the amount of speakers who choose $\theta$.$^{21}$ This announcement strategy will often induce pluralistic ignorance.

Noting that $\chi^i(\hat{G}_P)$ is the minimum informative context in game $\hat{G}_P$, we can now state our result. The proof is in Supplementary Materials 5.

**Result 6.** Suppose that $\pi \geq \hat{\pi} \geq (\beta + 1)/2\beta$ in game $\hat{G}_P$, where $\hat{\pi}$ is defined in Result 3. Further, suppose that social planner $P$ wants to maximize social welfare function $\hat{W}$, where $\hat{W}$ is either $W_1$ with $\gamma_H > 1$ and $\gamma_L > 1$, $W_2$ with $\beta \in (1, \hat{\beta})$ where $\hat{\beta}$ is defined in Result 5, or $W_3$. Then there exist $\Gamma_H(\chi, \pi, \gamma_L, \beta, I, \hat{W}) > 0$, $\Gamma_L(\chi, \pi, \gamma_H, \beta, I, \hat{W}) > 0$ and $\Gamma(\hat{G}_P, \hat{W}) \in (0, \min\{\Gamma_H, \Gamma_L\})$ such that,

1. If $|\gamma_H - \gamma_L| < \Gamma$, then

   • If the group $I$ is large enough, that is $I \geq I(\hat{G}_P)$ for some $I(\hat{G}_P) > 2$, then the social planner $P$ chooses an announcement strategy $\hat{A}$ that fully reveals the population.

   • If the group $I$ is small enough, that is $I \leq I(\hat{G}_P)$ for some $I(\hat{G}_P) < I(\hat{G}_P)$, and the context is close enough to 0.5, that is $\chi < \bar{\chi}$ for some $\bar{\chi} \in [0.5, \chi^i(\hat{G}_P)]$, then $P$ follows an uninformative announcement strategy: $P(\psi = 1 \mid h_1, \hat{A}) = P(\psi = 1 \mid h_1)$ for all $i, \hat{A} \in \{H, L\}$.

   • If $I \leq I(\hat{G}_P)$ and the context is informative, or $\chi \geq \chi^i(\hat{G}_P)$, then $P$ reveals that the population $\psi$ is equal to $H$ with announcement $\hat{A} \in \{H, L\}$, and increases the uncertainty about the context with announcement not-$\hat{A}$: $P(\psi = H \mid \text{not-}\hat{A}) \in [1 - \chi^i(\hat{G}_P), \chi^i(\hat{G}_P)]$.

2. If instead $\gamma_H - \gamma_L > \Gamma_H$ and the context is informative, then $P$ follows an uninformative announcement strategy. Furthermore, $\Gamma_H < \infty$.

3. If instead $\gamma_L - \gamma_H > \Gamma_L$ and the context $\chi < 1$ is informative, then $P$ reveals the population $\psi$ is $H$ with announcement $\hat{A} \in \{H, L\}$, and sends a message not-$\hat{A}$ that places beliefs over

---

$^{21}$Take an announcement strategy $A^*$ that maximizes the probability speakers choose $\theta$. Then $\Gamma_\theta(\chi, \pi, \gamma_{\text{not-}\theta}, \beta, I, \hat{W}) > 0$ is the minimum value of $\gamma_{\theta} - \gamma_{\text{not-}\theta}$ such that it is optimal for social planner $P$ to choose $A^*$.
the context outside of the informative range: \( P(\psi = H \mid not - \hat{A}) < \chi^i(\hat{G}_P). \) Furthermore, \( \Gamma_L < \infty. \)

In the introduction I claimed that a social planner should decrease second-order uncertainty to diminish pluralistic ignorance in a public opinion topic, such as attitudes towards segregation in the U.S. in the sixties. However, a University Dean should increase second-order uncertainty to get students to reveal their true attitudes in conversations. This is the logic of part 1 of Result 6, which describes the optimal announcement strategy when the difference in externalities from either action, or \( |\gamma_H - \gamma_L| \), is small. For large groups such as whites in the U.S. in the sixties, the social planner wants to reveal her private information about the majority true attitude, which she does by revealing the population from which the group is drawn. For small groups such as conversations among students, the social planner wants to increase the uncertainty about the majority true attitude, which she does by sending an announcement that updates the context to the range of non-informative contexts. This Result follows from how pluralistic ignorance varies with group size (Result 4) coupled with the social welfare benefits of diminishing pluralistic ignorance when externalities are similar for both actions (Result 5). In section 9.2 I discuss empirical evidence regarding information dissemination policies in large and small groups.

A social planner actually prefers pluralistic ignorance if it leads individuals to choose an action that has large positive externalities. The announcement strategies in the parts 2 and 3 of Result 6 reflect this motivation. To illustrate, suppose that a Dean in a university is worried about conversations that encourage drinking. Drinking leads to very negative externalities: accidents, sexual assault, lack of productivity. A social planner may then want to foster the idea that most students do not like to drink. Whether or not beliefs are accurate, having individuals expressing that det do not like to drink because they think their peers do not drink would be socially optimal. Pluralistic ignorance would then be acceptable to the Dean.\(^{22}\)

\(^{22}\)The model could itself be used as a stylized model of social drinking. Individuals take turns deciding whether or not to drink, and all actions are observed. Some people like drinking, and others don’t, and this is private information. Those who like drinking prefer that others drink, and those who do not like drinking dislike drunks. In addition to these material payoffs, individuals also care about how they expect to be judged by others, who are critical of others who do not share their taste regarding drinking. A Dean would want to encourage conversations that discourage drinking and
Although I have assumed that the Dean can perfectly commit *ex ante* to an announcement strategy, in practice these incentives lead to a problem of credibility. If observers of the Dean’s announcement know that the Dean would be willing to generate pluralistic ignorance, they may be suspicious of an announcement that claims the distribution of types is different from what they believe. This may explain the mixed evidence of these types of information interventions on reducing alcohol consumption in universities (see, e.g., Kenny et al., 2011). A full consideration of these concerns are outside the scope of this paper.

Result 6 does not pin down the exact optimal policy in several cases.\textsuperscript{23} Identifying the optimal announcement strategy is complicated because of two factors. First, to send a message $\hat{A}$ that increases uncertainty over context, there must be an alternative message $\text{not} - \hat{A}$ that decreases uncertainty. This creates a trade-off. Second, even when a non-informative context is optimal, moving the context away from 0.5 may be best. These complications make it hard to know what the optimal announcement strategy is in many cases. The proof in Supplementary Materials 5 discusses limitations to extending the Result further. Given the complications of maximizing a social welfare function, the social planner may choose instead to minimize pluralistic ignorance whenever externalities are not too asymmetric. This simpler objective is positively but imperfectly related to increasing social welfare. The following result states that the optimal announcement strategies for this alternative objective function are similar to those in Result 6.

**Corollary 1.** Suppose that $\pi \geq \hat{\pi} \geq (\beta + 1)/2\beta$ in game $\hat{G}_P$. Further, suppose that social planner $P$ wants to minimize pluralistic ignorance unless $\gamma_\theta - \gamma_{\text{not-}\theta} > \Gamma_\theta(\hat{G}_P)$, in which case $P$ wants to maximize pluralistic ignorance on $\theta \in \{H, L\}$. Then the announcement strategies from Result 6 are optimal, with possibly different thresholds for $I(\hat{G}_P), I(\hat{G}_P)$, and $\tilde{\chi}$.

\textsuperscript{23}For example, when there are small groups with small differences in externalities, it simply states that one of the announcements will update beliefs about context towards somewhere in the non-informative range. As a second example, the result is silent about the optimal policy when the externalities from action $H$ are larger than from action $L$ and the context is non-informative.
7.2 Co-opting First Movers

When the context is not informative, first movers’ actions have a large impact on others’ actions. A social planner may take advantage of this by co-opting first movers into choosing her own preferred type. Examples of this type of policy are not hard to find, from astroturfing (Bienkov, 2012, Cho et al., 2011, Monbiot, 2011) to interventions of Chinese government propaganda into social media feeds (King, Pan and Roberts, 2013).

Although I will not do so explicitly, this policy can be readily captured by a simple modification to the model. A social planner may pay a cost to ‘co-opt’ first movers into choosing the action the social planner is interested in promoting. Individuals observe a noisy public signal of whether each speaker was co-opted. The signal is informative: it is more likely to indicate that an individual was co-opted if the speaker was co-opted. The less uncertainty individuals have about how many first-movers the policymaker co-opted, the less effective these movers will be in swaying others’ actions. Individuals would not infer information about the group’s majority type from co-opted first movers’ action, since the type they are choosing is simply the preferred type of the social planner.

This simple modification to the setup yields a couple of predictions regarding when the policy is more effective, and therefore widespread. The first prediction is that co-opting is more effective and widespread when individuals in a group know each other less, or have fewer opportunities to interact with each other. These conditions would increase the uncertainty over how many individuals were co-opted. Co-opting should then be more effective among groups of strangers, or when the communication in a group is more rigid. Relatedly, I posit that small groups give more opportunities for closer interactions. This implies that co-opting will be more effective and widespread for large groups.

The second prediction is that co-opting first movers will be more effective and widespread when there is higher uncertainty over the distribution of types – that is, in non-informative contexts. This follows from the fact that individuals know that others will choose the same action in informative contexts. We should expect a higher investment in astroturfing or social media planting during times of social upheaval, when a longstanding norm is being questioned, or when a social planner
tries to influence a topic that is obscure to the public.

### 7.3 Varying the Weight on Social Expectations

A different approach to disseminating information is to foster the minority-attitude holders to be more willing to speak their mind. The Dean of a university may want students to express their true attitudes in conversations, or a policymaker may want truthful responses in surveys. The principal may then want to reduce whether speakers feel judged or whether they care about being judged – captured by decreasing the weight on social expectations $\beta$ in the model. This may be promoted by the Dean for small groups by creating safe spaces where students are encouraged to speak their minds freely. A policymaker worried about truthful survey reporting could lower $\beta$ by anonymizing responses, or with procedures such as list experiments, randomized responses, or endorsement experiments that allow survey respondents to partly conceal their attitude from the surveyor (Rosenfeld, Imai and Shapiro, 2016).\textsuperscript{24}

To think further about this policy, we must consider the question of what would happen if we relaxed the assumption that all individuals put the same weight on social expectations. The results would be qualitatively similar as long as $\beta$ is large enough, since speakers would still pool once they believed strongly enough that most were of a certain type. However, if some speakers have a low enough $\beta$, they would always choose their type since they do not care enough about social expectations. If all individuals were drawn independently from the same population, then for a large enough group there would never be pluralistic ignorance – an arbitrarily large number of speakers will reveal their type, and by the law of large numbers, individuals would eventually have an arbitrarily accurate estimate of the majority type of the group. However, this would not necessarily be the case if those with low $\beta$ were drawn from a different population, or with the same population but with a bias. This would allow a weaker inference from those with low $\beta$ to the majority type in the rest of the group. Intuitively, if there is a blunt subset of the group who always chooses their type, others will learn about the distribution of the whole group only if those

---

\textsuperscript{24}In section 9 I use the model to further think about empirical strategies for diminishing social expectations concerns.
who are blunt are a representative sample of the rest of the group, or a large enough proportion of
the group.

Let’s now apply this logic to the policy of decreasing the weight on social expectations $\beta$. The
success of the policy will then depend on those who are more willing to speak their minds as a
result of the policy. The policy will be more successful the more they are believed to be drawn
from the same population as everyone else, and the larger they are as a fraction of the group.

For reasons discussed in sections 6 and 7, a social planner may want to increase pluralistic
ignorance. In this case, a social planner may want to increase the weight on social expectations
$\beta$ to make speakers more willing to pool. The social planner may achieve this higher weight by
increasing the perception of the social consequences of an action.

8 Empirically Testing the Model

In this section I first consider how to empirically distinguish between the dynamics of non-informative
and informative contexts. Second, I consider work that directly tests concerns over expected judg-
ments. Third, I propose further testable predictions regarding information campaigns. Fourth, I
consider a laboratory approach to testing the small-group predictions of the model.

Distinguishing between dynamics. Results 1 and 3 show that there are two channels through
which stable beliefs over types may arise – that is, beliefs strong enough that individuals all choose
the same action. The first is an informative context. The second is a non-informative context
with a sufficiently large run of first movers choosing the same action – a herding mechanism.
In small groups, these are easy to distinguish empirically – there will be higher diversity in the
distribution of actions across and within groups with the herding mechanism. I am not aware of
any experimental tests of this prediction.

In large groups, the channels that lead to stable beliefs over types may seem observationally
equivalent at first glance. However, there are ways to distinguish them empirically. One approach
echoes the discussion in Eyster and Rabin (2010). When the stability in beliefs arises from herding,
individuals will stop choosing their type when there is just enough evidence that they would be judged better by parroting the perceived majority type. That is, once the revealed types are just enough for an individual to herd, nobody else reveals their type. In contrast, with an informative context, there can be an arbitrarily strong belief over the majority type of the population. To be more specific, for any precision $\pi$ large enough, the context $\chi$ can be set high enough that the strength of beliefs is stronger with an informative context than with herding. This provides one approach for distinguishing between the two types of stabilities in beliefs. Of course, this conclusion assumes that individuals are updating their beliefs as Bayesians, which may not be empirically accurate. Inasmuch as individuals are updating na"ively, for example by ignoring how the actions others have observed affects their own actions, then the herding mechanism may lead to arbitrarily strong beliefs.

A second approach to distinguish between the sources of stability in beliefs is to consider the information publicly available about the distribution of types. Stability comes from an informative context in a public opinion topic if widely available survey results are perceived to be truthfully reported, representative, and to show a clear majority. A non-informative context arises when survey data from a topic is not widely available, there has been a period of flux in attitudes, or cultural heterogeneity makes the distribution of attitudes uncertain. Against this backdrop, stability may come from a variety of mechanisms in which certain attitudes are heard prominently – particularly vocal social movements; campaigns to promote a viewpoint (such as political ads claiming an issue is what America is ‘really about’ or advertising campaigns promoting standards of taste in consumption); or prominent figures who espouse a vision of what they think is at the heart of a society.

In sum, using publicly available information across topics as a proxy for stability of beliefs allows us to test for the variation in pluralistic ignorance across informative and uninformative

---

25 Although truthful reporting of types can never be guaranteed, best practices in reputable surveys include anonymity, privacy and neutrally worded questions – consistent with respondents caring about how they think they will be judged by the surveyor, as in the model.

26 These examples raise the question of who gets to be the first mover in a conversation, which in the model is exogenous. One approach to endogenizing this order was discussed in in section 7.2.
contexts.\footnote{Shamir and Shamir (2000), who studied public opinion in Israel over 24 public opinion topics, in fact do find that topics that were more ‘publicly visible’ were less likely to have pluralistic ignorance. This is predicted by the model, since the authors are studying public views on a topic, which correspond to large groups in the model.} As mentioned in the introduction and in section 4, whether the communication is done in a private conversation or publicly affects the size of the group, so the main results can be tested. An additional testable implication that follows from the discussion is that publicly available information about a clear majority in a given topic decreases the likelihood of social movements, campaigns or social entrepreneurs.

**Testing concerns over expected judgments.** A different approach to testing the model is to test for concerns over expected judgments. Evidence from this comes from Bursztyn, Egorov and Fiorin (2017). In their study, they take advantage of Trump’s election. 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.

**Heterogeneous impact of providing information.** In section 7 I proposed that two variables affect the impact of providing information about the distribution of true attitudes: the externalities of an action and the perceived intentions of the social planners behind the information campaigns. That is, consider an information campaign that reveals a surprising distribution of attitudes on some topic. I suggested these campaigns less likely to work if there is a social planner that is perceived to want individuals to take an action supported by the surprising distribution. Further, the success should be even lower if the social planner cannot commit credibly to provide an unbiased estimate of the true distribution of attitudes. These incentive compatibility problems may explain the mixed results behind social information campaigns (e.g. Kenny et al., 2011).

**Laboratory tests of small group interactions.** There has been much experimental work on how individuals act in ways that do not match their private information when others disagree with them (I’ve mentioned examples above, and Asch, 1956, is a classic reference). To the best of my
knowledge the literature has focused on the impact of what others do or prefer on the behavior of an individual. We do not have much experimental evidence for the conditions under which pluralistic ignorance is more likely to arise as a consequence of a group interaction. Results 1, 2, 3 and 4 provide a theoretical starting point for experiments – testing the prediction that pluralistic ignorance is more likely to arise when the context is uninformative. The procedure would be relatively straightforward. Groups of small sizes can hold stylized conversations in a lab setting, and their true attitudes compared with their expressed attitudes.

9 Review of the Theoretical and Empirical Literature

In this section I conduct a more thorough review of the literature. The roadmap is as follows. In subsection 9.1 I argue that the features of pluralistic ignorance captured by the model have not been captured by past theoretical work, both formal and informal. In subsection 9.2 I discuss empirical evidence of pluralistic ignorance.

9.1 Alternative Approaches to Pluralistic Ignorance

I will argue that past formalizations do not capture the features of pluralistic ignorance from my model.

‘Pluralistic ignorance’ has received several definitions in both empirical and theoretical work. In this paper, I have focused on the original use of the phrase, as captured by Katz and Allport (1931) and O’Gorman (1975): ‘a majority of group members privately reject a norm, but incorrectly assume that most others accept it, and therefore go along with it.’ Alternative theoretical approaches to pluralistic ignorance do not capture this definition.

One approach to pluralistic ignorance is to suppose individuals want to take an action $a$ as long as enough others also take that action (so called strategic complementarities), but are misinformed about others’ threshold for action (e.g. Chwe, 2000, Kuran, 1997). A second approach to

\[ \text{Equation} \]
pluralistic ignorance comes from models where individuals are motivated to signal that their type matches what others commonly consider a desirable type (e.g. Bernheim, 1994, Bénabou and Tirole, 2011, Benabou and Tirole, 2012, Ellingsen, Johannesson et al., 2008).\textsuperscript{29} Models using both of these approaches have assumed common knowledge over the ‘good’ action. In the first approach, the highest equilibrium payoff is commonly known to come from choosing a specific action when enough others choose the same action. In the second approach, the type that provides most reputational benefits is exogenously given, and equilibrium reputational benefits increase the closer an action is to the ideal point of the optimal type.\textsuperscript{30} But then it is never the case that individuals will choose an action because they hold an incorrect belief of what the ‘good’ action is. But this is essential to the definition of pluralistic ignorance: it is precisely what is meant by ‘going along with the norm’ that is mistakenly thought to be accepted by others.

The closest setup to my model is found in Bursztyn, Egorov and Fiorin (2017) and Bursztyn, González and Yanagizawa-Drott (2018), who also assume individuals trade off choosing their type with signaling what they believe is the majority type. They assume a true distribution of types in a population, and a commonly known, exogenously given distribution that everyone believes is true. In this model, pluralistic ignorance and incorrect beliefs are assumed, not derived endogenously.

\subsection*{9.2 Empirical Evidence of Pluralistic Ignorance}

Here I consider the challenges and methodological solutions to estimating pluralistic ignorance.

As mentioned in the introduction, there are many studies documenting pluralistic ignorance across a variety of settings. The basic methodology for measuring pluralistic ignorance is to ask individuals their attitudes on some topic, ask them their beliefs over the distribution of attitudes on

\footnote{In particular, when those who are most enthusiastic about an activity interact most with others regarding that activity, individuals will oversample enthusiasm.}

\footnote{Although Bernheim (1994) allows for a more flexible reputational function, it is also exogenously given and commonly known. In my model, \textit{which} action yields perceived reputational benefit is endogenously determined.}

\footnote{Like in my model, Sliwka (2007) assumes that `conformists’ act according to whatever the majority type is. However, the model has only an informed principal and an agent, and the results focus on whether the principal reveals her information truthfully. Benabou and Tirole (2012) give a different but complementary definition of pluralistic ignorance within their model, which is that there is equilibrium misinformation about the intensity of preference over the commonly desirable type.}
the topic (typically, ask them about the majority attitude), and then compare the majority attitude estimated from the first question with the majority attitude estimated from the second. There are several concerns with this approach.

A first concern is that individuals may not report their true attitude. Surveys studying pluralistic ignorance typically study socially sensitive subjects, so there is a concern that they may be lying to the surveyor about their own attitude and telling the truth about what they believe the majority attitude is. It may well be that most whites in the U.S. in the 60’s were pro-segregationist, but they did not want to admit that to the surveyor. This problem is exacerbated by the fact that the surveys were done face to face. Best practice in survey methodology is to minimize these concerns by making subjects uncertain over how their answers would be judged, such as with careful wording of the questions, or through ensuring a non-judgmental appearance of the surveyors themselves. Another way to minimize these concerns is to make the surveyor care less about how they think they will be judged, such as through anonymous responses or providing plausible deniability. For example, list experiments ask subjects to report how many behaviors they’ve engaged in out of a list (Rosenfeld, Imai and Shapiro, 2016). The list includes a sensitive behavior of interest. By reporting the number of behaviors they’ve engaged in, subjects do not report whether they engage in the sensitive behavior directly, but the researcher can calculate the incidence in the population.

A second concern is that individuals may not be properly motivated to give their best guess as to the majority attitude. This concern is easier to solve, as individuals may be incentivized to guess the majority attitude of other survey respondents. However, this approach will not work if surveyors believe that others will distort their true attitude. Therefore, this approach only works if the first concern has been reasonably solved, and if surveyors understand this. This type of approach, coupled with anonymous surveying, was followed by Bursztyn, González and Yanagizawa-Drott (2018).

A third concern is that measured pluralistic ignorance may simply be noise. If pluralistic ignorance does matter, then we should expect that correcting beliefs should lead to a change in behavior.

---

31 Indeed, asking subjects what they think others think or do is a survey technique sometimes used to ask sensitive questions, such as on the prevalence of corruption (e.g. Mauro, 1995).
Wenzel (2005) and Bursztyn, González and Yanagizawa-Drott (2018) provide evidence to that effect, as does the wider literature on campaigns that publicize average behavior (see Kenny et al., 2011, for a review). Wenzel (2005) found that individuals overestimate the average acceptance of tax evasion. He finds that informing individuals of their misperception increases tax compliance in the field as well as with hypothetical questions. Bursztyn, González and Yanagizawa-Drott (2018) randomly informed Saudi men that others in their social network were more supportive of female labor force participation than they originally thought. Treated subjects were more likely to sign their wives up for a job interview, and to later apply for a job outside of home. These effects were concentrated among those who overestimated the amount of agreement on limiting female labor force participation.

It is worth noting that correcting misinformation may have an opposite effect than that predicted by the model, due to the presence of other forces. Cantoni et al. (Forthcoming) found that an individual who underestimates others’ intentions to protest is less likely to protest herself if she is provided information of their intentions. One explanation the authors provide for this decrease is that protesting was seen as a threshold public good problem, in which there was no need to protest if enough others portested.

10 Conclusion

In this paper I introduced a model of dynamic attitude expression with which I captured pluralistic ignorance as an equilibrium outcome. The theoretical contributions of the paper are fourfold. First, it captures what I’ve argued are key features of pluralistic ignorance: most individuals act reluctantly, most believe most others do not act reluctantly, and it is the underestimation of others’ reluctance that motivates an individual’s reluctance. Second, the model provides a novel 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 types are drawn, and lowest in large groups when there is low uncertainty. Third, the model provides a
link between pluralistic ignorance and collective action problems: a perverse failure of collective action can arise in equilibrium, where most individuals avoid choosing the action that they would prefer in the absence of social expectations, the action they do choose minimizes externalities, and most individuals judge others negatively for the actions they choose. Fourth, the paper shows how optimal information dissemination policies depend on group size: a social planner interested in maximizing social welfare or minimizing pluralistic ignorance will provide information to decrease uncertainty about the distribution of types in large groups, but provide information to increase uncertainty in small groups. Other policies are also considered, such as co-opting first movers when there is initial uncertainty over the distribution of types. The paper relates these results to the empirical evidence, and proposes methods of measuring and testing the model’s predictions.

Possible next steps include the following: allowing individuals to remain silent instead of expressing an attitude; considering biased social updating; an experiment which tests for the emergence of pluralistic ignorance in small-group dynamics; a study of the determinants of pluralistic ignorance across public opinion topics; and a theory of information dissemination policies where the social planner cannot perfectly commit ex ante to an announcement strategy. Other directions include concrete political economy applications such as the use of censorship in small-group interactions to affect public opinion and the types of leaders that can break a group out of pluralistic ignorance.

CIDE, Mexico City

A Proof of Result 1, and a Justification of the Equilibrium Concept

In order to establish Result 1, I will prove a more general result for a weaker solution concept. I will end this section with a discussion of how the more general result justifies the focus on the solution concept in Definition 1. The proof of all Lemmas except Lemma 2 are in the Supplementary
Materials.

Lemma 2.  
1. If the context is informative, all speakers pooling on \( H \) is a PBE strategy.

2. If the context is uninformative or semi-informative, speaker 1 separates.

3. Suppose that at history \( h_i \), (a) speaker \( i \) of type \( L \) expects others to be most likely of type \( L \), or \( \mathbb{E}_{j \neq i}(P(\theta_j = L \mid h_i, \theta_i = L)) > 1/2 \); and (b) a judge \( j \) without private information believes the speaker \( i \) is most likely of type \( H \), or \( P(\theta_i = H \mid h_i) > 1/2 \). Then separating is a PBE strategy if \( -\Delta(h_i)/(I - 1) < 1/\beta \). Conditions (a) and (b) only happen on the equilibrium path in a non-reverse PBE when the context is semi-informative. Further, the conditions imply \( -\Delta(h_i) > 0 \).

Suppose that along the equilibrium path of play, all speakers \( k < i \) have separated. Further suppose that at history \( h_i \), if (a) and (b) hold, then \( -\Delta(h_i)/(I - 1) < 1/\beta \) holds.

4. There is some threshold \( n(a, h_i; G) > 0 \) such that if the lead for action \( a \in \{H, L\} \) is equal to \( n(a, h_i; G) \), pooling on \( a \) is a PBE strategy for speaker \( i \) and all speakers \( l > i \). Otherwise, separating is a PBE strategy for \( i \).

5. The threshold \( n(a, h_i; G) \) is finite if \( \pi > (1 + \beta)/(2\beta) \), and weakly decreasing in the weight on social expectations \( \beta \) and in the precision \( \pi \). For a fixed \( \Delta(h_i) \), the threshold weakly increases in \( i \). Further, \( n(H, i; G) \) and \( n(L, i; G) \) decrease and increase, respectively, in \( \chi \), with \( n(H, i; G) = n(L, i; G) \) for \( \chi = 1/2 \). When the context is informative, \( n(H, h_i; G) = 0 < n(L, h_i; G) \).

6. If \( \beta \in (1, 2) \), the pooling strategies described above are unique PBE strategies. If \( \beta > 2 \) and pooling is a non-reverse PBE strategy, separating is not. If for \( \beta > 1 \) separating is a non-reverse PBE strategy, pooling is not.

Proof. In order to establish the proof, I will first work my way towards establishing the following:
Claim 1. Individual $i$ of type $x$ deviates from a separating strategy if and only if:

$$
E_{j \neq i} \left( P(\theta_j = x \mid h_i, \theta_i = x) \mid h_i \right) < \frac{\beta - 1}{2\beta}
$$

(5)

If this condition holds, pooling on not-$x$ is a PBE strategy if the context is uninformative. If furthermore $\beta \in (1, 2)$, the pooling strategy is the unique PBE strategy.

Inequality (5) follows from considering the utility of type $\theta_i = x$ under a separating strategy. Recall that judge $j$ of type $x$ believes types match if she thinks it is more than 50% likely that $i$’s type is the same as hers, or $P(\theta_i = x \mid h_i, \theta_j = x) > 1/2$. If $i$ separates, only judges of type $x$ believe types match if $i$ chooses $x$. Therefore, speaker of type $x$ compares the social expectation from choosing $x$ and having only judges of type $x$ believe types match, or choosing not-$x$ and having only judges of type not-$x$ believe types match. This leads to inequality (5).

If speaker $i$ follows a non-reverse PBE strategy to pool on $x$, speaker $i + 1$ will have the same information about the group that speaker $i$ did. Speaker $i + 1$ will then face the same incentives as speaker $i$, so it will also be a non-reverse PBE strategy for $i + 1$ to pool on $x$. In fact, if (5) holds at the beginning of the game (when $i$ is equal to 1) and $\beta \in (1, 2)$, we can conclude from Claim 1 that all speakers will pool on $x$. For an uninformative context under these conditions, we just need to consider the history of play that leads a speaker to deviate from separating, knowing that pooling is the unique non-reverse PBE for all future speakers. Claim 1 and its proof will also help us analyze contexts which are semi-informative, as well as the case where $\beta > 2$.

To calculate an individual’s social expectations, we need to know how a speaker forms his beliefs over how he is judged by others. A individual $j$ who has already made a decision will continue to judge others, but may have revealed her type with her decision. Therefore, speaker $i$’s beliefs about how $j$ will judge him will depend on whether $j$ has separated. Let $\theta^o_j : \mathcal{H} \to \{-1, 0, +1\}$ be $j$’s observed type, a map from history $h_i$ to the value $-1$ if $j$ has revealed type $L$, to the value $+1$ if $j$ has revealed type $H$, and to the value $0$ if $j$ is a withholder.

When it is useful, I will explicitly condition $j$’s belief on her observed type $\theta^o_j$ – for example,
I will write \( P(\theta_i = H \mid h_i, \theta_j, \theta^o_j) \). Although the conditioning on \( \theta^o_j \) 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 speakers’ decisions will depend on their beliefs. When speaker \( i \) pools, judge \( j \)’s judgment of \( i \) will depend on her beliefs over \( i \)’s most likely type. In turn, speaker \( i \)’s decision will depend on his belief over the distribution of judges’ types. To capture what an individual \( k \) knows about the group, let the observed type lead be \( \Delta^o(h_i, \theta_k) \equiv \sum_{l \neq k} \theta^o_l(h_i) + (2 \times 1 \{ \theta_k = H \}) - 1 \), or the difference between the number of type \( H \) and type \( L \) signals observed by individual \( k \) of type \( \theta_k \) at history \( h_i \).

**Lemma 3.** Suppose at history \( h_i \), individuals are either revealers or withholders.

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

\[
f_j(z, \theta_j, \theta^o_j, \sum_{l \neq j} \theta^o_l) \equiv \left( \frac{\pi}{1 - \pi} \right)^{\Delta^o} [\pi - z] - \frac{1 - \chi}{\chi} [z - (1 - \pi)] > 0
\]

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

\[
g_i(z, \theta_i, \sum_{l \neq i} \theta^o_l) \equiv \left( \frac{\pi}{1 - \pi} \right)^{\Delta^o} \left[ \mathbb{E}_{j \neq i}(P(\theta_j = H \mid h_i, \psi = H)) - z \right] - \\
\frac{1 - \chi}{\chi} [z - \mathbb{E}_{j \neq i}(P(\theta_j = H \mid h_i, \psi = L))] > 0
\]

Expressions \( f_j > 0 \) and \( g_i > 0 \) in Lemma 3 characterize the beliefs that shape speakers’ and judges’ choices, in terms of primitives and the observed type lead. The first multiplicand of \( f_j \) and of \( g_i \) is the likelihood ratio of the precision \( \pi \), to the power of the observed type lead \( \Delta^o \). This term captures the information an individual has about the population based on the signals she or he has observed. Expressions \( f_j \) and \( g_i \) also depend positively on the context \( \chi \), which again is the prior probability that the population is \( \psi = H \). The expressions only differ in the terms in square
brackets. Note that the observed type is not specified for \(g_i\) since speakers are always withholders when they are making their decision.

For reasons I will explain shortly, we’ll focus on probabilities greater than \(1/2\). Therefore, let us adopt the following notational shorthand: 
\[
\hat{f}_j(\theta_j, \theta_{ij}^o, \sum_{l \neq j} \theta_l^o) \equiv f_j(1/2, \theta_j, \theta_{ij}^o, \sum_{l \neq j} \theta_l^o) \quad \text{and} \quad \hat{g}_i(\theta_i, \sum_{l \neq i} \theta_l^o) \equiv g_i(1/2, \theta_i, \sum_{l \neq i} \theta_l^o).
\]

**Lemma 4.** 1. The terms \(f_k\) and \(g_k\) increase in \(\theta_{ik}^o\), \(\sum_{l \neq k} \theta_l^o\), and \(\chi\), holding everything else constant.

2. Take \(\theta \in \{H, L\}\). Suppose \(\hat{f}_j(\theta, 0, 0) < 0\) and speakers before period \(i\) have all separated. Then \(\hat{g}_i(\theta_i, x) < 0\) for \(x \leq 0\); \(\hat{g}_i(\theta_i, x) > \hat{f}_j(\theta, 0, x)\) for \(x > 0\) and \(\hat{f}_j(\theta, 0, x) < 0\); and \(\hat{g}_i(\theta_i, x) > 0\) for \(x > 0\) and \(\hat{f}_j(\theta, 0, x) > 0\). For a fixed \(x\), \(|\hat{g}_i(\theta_i, x)|\) decreases in \(i\) unless \(x > 0\) and \(\hat{f}_j(\theta, 0, x) < 0\), in which case \(\hat{g}_i(\theta_i, x)\) increases in \(i\). The result holds with all signs reversed, and ‘increases’ changed to ‘decreases’ in the last sentence.

3. The terms \(\hat{f}_k(\theta, \cdot, x)\) and \(\hat{g}_k(\theta, \cdot, x)\) increase in \(\pi\) if \(\Delta(h_i) > 0\) and decrease in \(\pi\) if \(\Delta(h_i) < 0\).

4. The context is uninformative if and only if \(\hat{g}_1(L, 0, 0) < 0 < \hat{g}_1(H, 0, 0)\). The context is informative if and only if \(g_1((\beta - 1)/2\beta, 0, 0, 0) < 0\), or equivalently, (5) holds for \(i = 1\).

By Lemma 4, a speaker of type \(H\) believes more of his judges are of type \(H\) than does a speaker of type \(L\): 
\[
\mathbb{E}_{j \neq i}(P(\theta_j = H \mid h_i, \theta_i = H)) \text{ is larger than } \mathbb{E}_{j \neq i}(P(\theta_j = H \mid h_i, \theta_i = L)).
\]
Each type may believe most judges are of the same type as himself. In this case, each type gets higher social expectations from revealing his type than from revealing the opposite type, or \(\hat{g}_i(L, x) < 0 < \hat{g}_i(H, x)\), so separating is a PBE strategy. Alternatively, both types may believe most judges are of type \(x\).

Consider the case where \(\beta \in (1, 2)\). An individual chooses his type \(x \in \{H, L\}\) if the differ-
ence in social expectation from choosing not-$x$ versus choosing $x$ is less than $1/2$:

$$\frac{1}{|I| - 1} \sum_{j \neq i} \mathbb{E}(J_{j,i}(\text{not} - x) \mid h_i, \theta_i) - \frac{1}{|I| - 1} \sum_{j \neq i} \mathbb{E}(J_{j,i}(x) \mid h_i, \theta_i) < \frac{1}{\beta} \in \left(\frac{1}{2}, 1\right)$$

The focus of Lemma 4 on probabilities greater than $1/2$ can now be seen more clearly. Judges care about the probability that the speaker is of type $H$ with probability greater than $1/2$ (or $\hat{f}_j(\theta_j, \sum_{l \neq j} \theta_l) > 0$) because that determines whether they believe the speaker’s type matches their own. Speakers care about whether judges are on average most likely of type $H$ (or $\hat{g}_i(\theta_i, \sum_{l \neq i} \theta_l) > 0$) because it is a sufficient condition for them to be willing to choose a type different from their own.

**Lemma 5.**

1. Suppose that at history $h_i$, speakers have either separated or pooled. Then it is not a PBE strategy for both types of speaker $i$ to mix.

2. Suppose that $\beta \in (1,2)$, that both types of speaker $i$ believe most judges are of type $a \in \{H, L\}$, and that both revealer and withholder judges of type $a$ believe withholders are most likely of type $a$. Then pooling or semi-pooling on not-$a$ are not PBE strategies.

3. Suppose $\hat{f}_j(L, -1, x) < 0$ and $\hat{g}_i(L, x) < 0$ for some integer $x$. Then either (a) pooling on $L$ is a PBE strategy, while separating and semi-pooling on $L$ are not, or (b) separating is a PBE strategy. The analogous result holds for $\hat{f}_j(H, 1, x) > 0$ and $\hat{g}_i(H, x) > 0$.

Lemma 5 will allow us pin down a unique PBE strategy.

I now close the argument by describing the PBE dynamics. Suppose first that there is an equal number of publicly observed signals of type $L$ and of type $H$, or $\sum_{k \neq i} \theta_k = 0$. I will show that in this scenario either it is a unique PBE strategy for the first speaker to pool (and consequently for all speakers to pool), or separating is always a PBE strategy. Therefore, the only outcome where all speakers pool when $\sum_{k \neq i} \theta_k = 0$ is determined completely by the environment: the context $\chi$ and the precision $\pi$. The conditions that lead all speakers to pool are the conditions that define an informative context.
By Lemma 4, it follows that \( \hat{f}_j(\theta, 0, 0) \) has the same sign as \( \hat{g}_i(\theta, 0) \) for any \( \theta \in \{H, L\} \). Separating is a PBE strategy in an uninformative context, where \( \hat{g}_i(L, 0) < 0 < \hat{g}_i(H, 0) \), which holds for all \( i \) if it holds for \( i = 1 \). Note that this is not necessarily a unique non-reverse PBE strategy.

Suppose the context is not uninformative: only type \( H \) judges would believe types match if speaker \( i \) pools, and both types of a speaker would believe most judges are of type \( H \) (again by Lemma 4, and recalling the assumption that \( \chi \geq 1/2 \)). Since the social expectations from pooling on \( H \) and from revealing \( H \) are the same, there is no semi-pooling on \( H \). Strategies where both types mix are ruled out by Lemma 5, and either there is a separating PBE strategy or pooling on \( H \) is a PBE strategy (as shown in the proof of Lemma 5). Furthermore by Lemma 5, pooling or semi-pooling on \( L \) are not PBE strategies for \( \beta \in (1, 2) \). Therefore, when \( x = 0 \), the context is not uninformative and \( \beta \in (1, 2) \), the non-reverse PBE strategy is unique.

If the context is informative, pooling on \( H \) is a PBE strategy for speaker 1 – and therefore it is a PBE for all speakers to pool on \( H \). If the context is semi-informative, separating is a PBE strategy for speaker 1. By Lemma 4, \( |\hat{g}_i(\theta, 0)| \) decreases with \( i \) for any \( \theta \in \{H, L\} \). The more there are a balanced number of type \( H \) and type \( L \) speakers who have revealed their type, the smaller the incentives to pool on \( H \), and the higher the incentives to separate. Therefore, speakers will continue to separate for all periods with an observed type lead of 0 (again using Lemma 5).

Since we’ve assumed without loss of generality that \( \chi \geq 1/2 \), it follows from Lemma 3 that \( \hat{f}_j(H, 0, 0) > 0 \). Then if \( x > 0 \), \( \hat{f}_j(H, 0, x) > \hat{f}_j(H, 1, x) \geq \hat{f}_j(H, 1, 1) = \hat{f}_j(H, 0, 0) > 0 \) and \( \hat{g}_i(H, x) > 0 \) by Lemma 4. Then judges of type \( H \) believe types match if speaker \( i \) pools, and speaker \( i \) of type \( H \) believes most judges are of type \( H \). Since the social expectations from pooling on \( H \) are at least as high as the social expectations from revealing \( H \), there is either a separating PBE strategy or pooling on \( H \) is a PBE strategy. This uses Lemma 5. From Lemma 5 it also follows that the PBE strategies are the unique non-reverse PBE strategies if \( \beta \in (1, 2) \). As \( x \) increases, the incentives for speaker \( i \) of type \( L \) to deviate from a separating strategy increase.

Now suppose that \( x < 0 \). We’ll again apply Lemmas 4 and 5, as well as reasoning analogous
to the above paragraphs. Separating is a PBE strategy if $\hat{g}_i(L, x) < 0 < \hat{g}_i(H, x)$. Note again that this need not be the unique non-reverse PBE strategy. There is either a separating PBE strategy or pooling on $H$ is a PBE strategy if $\hat{f}_j(H, 0, x) < 0$ and $\hat{g}_i(H, x) < 0$. These PBE strategies are the unique non-reverse PBE strategies if $\beta \in (1, 2)$. The incentives for speaker $i$ of type $H$ to deviate from a separating strategy increases as $x$ becomes more negative. Further, there are no other cases to consider when the context is uninformative and $x < 0$: using Lemma 4, $\hat{f}_j(H, 0, -1) = \hat{f}_j(H, 1, 0) = \hat{f}_j(H, 0, 0) \geq \hat{f}_j(H, 0, \hat{x})$ for $\hat{x} \leq -2$, $\hat{f}_j(H, 0, -1) > \hat{g}_i(H, -1) > 0$, and $\hat{g}_i(H, \hat{x}) < 0$.

Up to here is the proof of Claim 1, although we have covered more ground that allows us to consider semi-informative contexts.

Suppose that $\hat{f}_j(L, -1, x) > 0$ and $\hat{g}_i(H, x) < 0$. This can only happen on the equilibrium path if the context is semi-informative, and by the next-to-last paragraph, that $x < 0$. Speaker of type $H$ may deviate from a separating strategy since he believes most judges are of type $L$. If the speaker pools, revealers of type $L$ do not believe types match. The judges who believe types match are revealers of type $H$, withholders of type $H$, and in some cases withholders of type $L$ (Lemma 4). If the intuitive criterion can be applied, the out-of-equilibrium belief from a deviation of pooling on $\theta \in \{H, L\}$ would be that type not-$\theta$ reveals his type. But it may then be that either type $\theta$ would prefer to choose $\theta$ and reveal his type than to pool on not-$\theta$: this is easiest to see if only type $H$ judges believe types match after the speaker pools.

Then a pure PBE strategy is not guaranteed for some (perhaps the unique) out-of-equilibrium belief: speaker of type $H$ may deviate from choosing $H$ in a separating strategy, and speaker of type $\theta$ would deviate from pooling on not-$\theta$. In order for speaker of type $H$ to not deviate from a separating strategy, it must be the case that

$$P\left((\theta_j, \theta_j^\circ) = (L, 0) \mid h_i, \theta_i = H\right) + P\left((\theta_j, \theta_j^\circ) = (L, -1) \mid h_i, \theta_i = H\right)$$

$$-P\left((\theta_j, \theta_j^\circ) = (H, 0) \mid h_i, \theta_i = H\right) - P\left((\theta_j, \theta_j^\circ) = (H, 1) \mid h_i, \theta_i = H\right) < 1/\beta$$

where, for a given $i$ and $x < 0$, the left hand side of the inequality is maximized when there are no
revealers of type $H$, and when $P((\theta_j, \theta_i^o) = (L, 0) \mid h_i, \theta_i = H)$ is equal to $P((\theta_j, \theta_i^o) = (H, 0) \mid h_i, \theta_i = H)$, since at history $h_i, P((\theta_j, \theta_i^o) = (L, 0) \mid h_i, \theta_i = H) \geq P((\theta_j, \theta_i^o) = (H, 0) \mid h_i, \theta_i = H)$ – speaker of type $H$ believes it is more likely that a withholder is of type $L$ than of type $H$. This follows from the fact that $\hat{f}_j(H, 0, x) > \hat{f}_j(L, -1, x) > 0$, recalling that $\hat{f}_j(H, 0, x) > 0$ captures that a withholder of type $H$ believes a withholder is most likely of type $H$ (Lemma 3). Taking the maximum value of the left hand side yields $-\Delta(h_i)/(I - 1) < 1/\beta$.

The fifth paragraph of the statement of Lemma 2 refers to a threshold $n(a, h_i; G)$ that determines whether speaker $i$ pools on $a$. The proof of the comparative statics on $n$ follow from arguments presented above – Claim 1, Lemmas 3 and 4, and the equilibrium dynamics analysis of the preceding paragraphs. The finiteness of the threshold follows from Lemma 6 in appendix B. 

\[ \square \]

A.1 A Justification of the Narrowly Defined Solution Concept

Lemma 2 indicates that multiple non-reverse PBE strategies may exist for a speaker in a variety of situations – when separating is a non-reverse PBE strategy,\(^{32}\) or when the weight on social expectations $\beta$ is high enough. Furthermore, in semi-informative contexts, a non-reverse PBE strategy may not be pure at some histories. Lemma 2 provides conditions under which there exist pure non-reverse PBE strategies in semi-informative contexts, which I focus on in the main results.

In this section I use Lemma 2 to justify the use of a more narrowly defined solution concept than non-reverse PBE strategies, the equilibrium strategies from Definition 1.

\(^{32}\)Separating strategies are often the unique non-reverse PBE strategies. The exception arises when $\hat{f}_j(L, 0, x) < 0 < \hat{f}_j(H, 0, x)$ for some integer $x$. Then a semi-separating strategy can make revealer judges indifferent in their judgment. Therefore, type $\theta$ may not want to deviate from a separating strategy, but may prefer to have all judges believe types match than to choose $\theta$ and have only $\theta$ judges believe types match. Then a semi-separating strategy could also be sustained.
ficient condition so that non-reverse pure PBE strategies always exist, which I assume throughout the body of the text.

When $\beta > 2$, if a separating strategy is a non-reverse PBE strategy, it is the unique equilibrium strategy. Suppose pooling is a non-reverse PBE strategy and it leads judges of type $a$ to believe types match. Then either pooling on $a$ is the unique equilibrium strategy, or the only other non-reverse PBE strategies are to pool or semi-pool on not-$a$. This implies that pooling is the only type of equilibrium pure strategy. That the speaker would pool on any action is simply a consequence of the high weight on social expectations $\beta$.

The semi-pooling equilibrium on not-$a$ comes from the fact the action not-$a$ may support beliefs such that judges of type not-$a$ believe types match, revealer judges of type $a$ believe types do not match, and withholder judges of type $a$ randomize their judgment. Since only judges of type $a$ believe types match when the action is $a$, social expectations from choosing not-$a$ may be equalized given the randomization of withholder judges of type $a$.

There are then some histories where there is a pure PBE strategy of pooling on $a$ and a mixed PBE strategy of semi-pooling on not-$a$. The mixed PBE strategy is more informative. However, I will focus on the equilibria such that at these histories speakers pool on $a$. I provide three reasons. First, pure strategies are more tractable. Second, the equilibrium strategy I focus on is comparable to the one in the case of $\beta \in (1, 2)$, providing continuity in the analysis. Third, it is natural to look at equilibria where, if enough speakers reveal that they are of type $a$, the next speaker reacts by choosing $a$ with a higher probability.

### B Proof of Result 2

In this section I will present a stronger version of Result 2 that I will refer to as Lemma 7. I will present the arguments first, and then collect them in a statement of the Lemma.

In uninformative and semi-informative contexts there are two opposing forces affecting the probability of pluralistic ignorance as $I$ increases. First suppose that as $I$ grows, there is no change
in \( n(a, h_i; G) \) – which, recalling from Result 1, is the threshold that needs to be reached for all future speakers to pool on action \( a \). Then the first \( I' \) speakers follow the same strategy in a game of size \( I' \) as they do in a game of size \( I'' > I' \). But as \( I \) grows, the probability that most individuals are of a different type than what they are pooling on increases – there will be a higher proportion of individuals whose type is not-a with a probability between zero and one, and a lower proportion whose type is \( a \) with certainty.

On the other hand, and in contrast to the stated assumption, an increase in \( I \) weakly increases the threshold \( n(a, h_i; G) \) for \( a \in \{H, L\} \). To see this, consider the conditions for speaker \( i < I \) of type \( \theta_i = \hat{\theta} \in \{H, L\} \) to pool given a history where past speakers have chosen pure strategies.\(^{33}\) We showed in Result 1 that if a speaker pools in equilibrium, only one type of judge believes type match. Result 1 also pins down out-of-equilibrium beliefs using the intuitive criterion, so the condition can be written as follows:

\[
\frac{I - x - y - 1}{I - 1} \left( 2P(\theta_I = 1 - \hat{\theta} | h_i, \theta_i = \hat{\theta}) - 1 \right) + \frac{x - y}{I - 1} > \frac{1}{\beta}
\]  

(6)

where \( x \) is the number of speakers who have revealed type \( 1 - \hat{\theta} \) and \( y \) is the number of speakers who have revealed type \( \hat{\theta} \). As \( I \) grows, more weight is put on the term in parentheses than on the second summand. Further, the term in parentheses does not depend on \( I \). Since the term in parenthesis is between \(-1\) and \(1\), putting less weight on this term implies that a change in \( x \) or \( y \) has a smaller impact on the left hand side of the inequality. To illustrate, recall from section 4.2 that there is no separating for second movers when \( I = 2 \), but there may be when \( I = 3 \). The inequality also shows that it is easier to reach the threshold for pooling as the weight on social expectations \( \beta \) increases, since the left hand side does not depend on \( \beta \) and the right hand side decreases in \( \beta \).

An increase in the incentives for separating caused by an increase in \( I \) may decrease the probability of pluralistic ignorance because more individuals reveal their type. However, this effect is

\(^{33}\)Inequality (6) does not apply to \( i = I \). If \( i = I \), then by Result 1 either all speakers would have revealed their type, or there is some \( j < I \) such that all speakers \( k \in \{j, ... I\} \) pool.
Lemma 6. Suppose $\pi > (\beta+1)/2\beta$ and that the context is either uninformative or semi-informative with pure strategies being chosen up to history $h_i$. For a given $\chi$, $\pi$ and $\beta$, the maximum value of $n(a, h_i; G)$ for some $a \in \{H, L\}$ is given by:

$$n_{\text{max}}^\ast(h_i; G) \equiv \left\lceil \frac{\ln \left\{ \frac{1-\chi}{\chi} \left( \frac{\beta+1}{2\beta} - (1 - \pi) \right) \right\} - \ln \left\{ \pi - \frac{\beta+1}{2\beta} \right\}}{\ln \left\{ \frac{\pi}{1-\pi} \right\}} \right\rceil \in (0, \infty)$$

where $\lceil \cdot \rceil$ is the ceiling function. If $\pi \leq (\beta+1)/2\beta$, then for any value of $\chi \in [0.5, 1]$ $\lim_{I \to \infty} n_{\text{max}}^\ast = \infty$.

As we saw in section 4.2, the probability of pluralistic ignorance is zero when $I = 2$ and the context is uninformative or semi-informative. When $\pi > (\beta + 1)/2\beta$, we can use Lemma 6 to conclude that, for a large enough $I$, the probability of pluralistic ignorance is positive for all uninformative and semi-informative contexts. Indeed, in section 4.2 we saw that it is strictly positive when $I \to \infty$. Call $I + 1$ the minimum $I$ that leads to a positive probability of pluralistic ignorance in all the range of uninformative and semi-informative contexts. For $I \in [2, I]$, the range of contexts such that the probability of pluralistic ignorance is zero goes from 0.5 to a value $\hat{\chi}(I)$ — as $\chi$ increases, the threshold for pooling on $H$ becomes smaller than half of the group size (by Result 1). The function $\hat{\chi}(I)$ decreases in $I$ except for at most the number of increments of $I$ that increase the threshold for pooling — a logic similar to the one behind Lemma 6. Similarly, $\hat{\chi}(I)$ decreases in $\beta$ (again, by Result 1).

Let $\theta^n_I$ be the random variable representing the majority type of a group of size $I$. In informative contexts, the probability of pluralistic ignorance is equal to

$$\chi P(\theta^n_I = L | \psi = H) + (1 - \chi)P(\theta^n_I = L | \psi = L) \quad (7)$$

First, note that since $P(\theta^n_I = L | \psi = H) < P(\theta^n_I = L | \psi = L)$, the probability of pluralistic ignorance diminishes with $\chi$. Second, note that by the law of large numbers $P(\theta^n_I = L | \psi = H)$
diminishes with $I$ and tends to zero, while $P(\theta^m_i = L \mid \psi = L)$ increases with $I$ and tends to 1 (Casella and Berger, 2002). Therefore, the probability (7) goes to $1 - \chi$ as $I \to \infty$. The probability of pluralistic ignorance in an informative context can be made arbitrarily close to $\chi$ by making $I$ arbitrarily large. As an immediate corollary, the probability of pluralistic ignorance can be made as close to zero, or even zero, by making $I$ and $\chi$ large (again, assuming $\pi < (\beta + 1)/2\beta$). It is not necessarily the case that (7) decreases monotonically with $I$, however. This uses the fact that $P(\theta^m_i = L \mid \psi = L)$ increases in $I$. If an informative context $\chi$ is low enough, the probability of pluralistic ignorance when $I = 2$ (equal to $\chi(1 - \pi)^2 + (1 - \chi)\pi^2$) may be lower than the probability when $I \to \infty$ (equal to $1 - \chi$). This would happen if $\pi = 0.65$, $\chi = 0.8$ and $\beta = 10$. Notice the large value of $\beta$, which allows for the informative context to have a relatively wide range (i.e., it makes $\chi^i$ small). Despite these cases, the probability of pluralistic ignorance is increasing in $I$ for $\chi$ close enough to 1. This can be seen since, when $\chi = 1$, the increasing relationship holds for all parameter values.

Now consider the case where $\pi \leq (\beta + 1)/2\beta$. What this condition says is that the first speaker of type $\hat{\theta}$ would not want to follow a strategy of pooling on not-$\hat{\theta}$ even if he knew for certain that the population was not-$\hat{\theta}$. This is why there are no informative contexts in this case, as per Lemma 1. Further, as can be seen from (6), the incentives for any speaker tend to those of the first speaker as $I \to \infty$. Therefore, the threshold for pooling on either action goes to infinity as $I$ grows arbitrarily (as per Lemma 6). The probability of pluralistic ignorance is then equal to zero for very high and very low values of $I$ – for very high values of $I$ all speakers separate. There may be pluralistic ignorance for intermediate values, however. For $I = 5$, for instance, consider a realization of types such that the first two speakers reveal they are of type $H$. If the third individual is of type $L$, he will have one excess signal of type $H$. Therefore, he will know the first two individuals are of type $H$ for sure, and that the fourth and fifth are most likely of type $H$. This may provide a higher incentive to pool on $H$ than if he did not know anybody’s type but was sure the population was $\psi = H$.

I collect the above discussion as the main result.
Lemma 7. For a game $G$, suppose that $\pi > (\beta + 1)/2\beta$.

- For $\chi$ in the range of uninformative or semi-informative contexts, the probability of pluralistic increases monotonically in $I$ except for no more than $2 \times n^{\max}$ times. Further, there is a $I \geq 2$, weakly increasing in $\beta$, such that
  
  - The probability of pluralistic ignorance for a group of size $I$ is zero for $\chi = 0.5$.
  
  - The probability of pluralistic ignorance for a group of size $I + 1$ is positive for all $\chi$.
  
  - There exists a function $\hat{\chi}(I)$ such that, for a group of size $I \leq \hat{\chi}(I)$, the probability of pluralistic ignorance is zero for $\chi < \hat{\chi}(I) \leq \chi^i$, with $\hat{\chi}(I)$ weakly increasing as $\beta$ decreases. $\hat{\chi}(2) = \chi^i$.

- Suppose $\chi$ is in the range of informative contexts.
  
  - For any $\varepsilon > 0$ there exists a value $I(\varepsilon) > I$ such that the probability of pluralistic ignorance in a group of size $I \geq I(\varepsilon)$ is less than $1 - \chi + \varepsilon$.
  
  - There exists some $\hat{\chi} \in [\chi^i, 1]$ such that for $\chi > \hat{\chi}$, the probability of pluralistic ignorance decreases monotonically with $I$.
  
  - The probability of pluralistic ignorance in an informative context diminishes with $\chi$ in the range $[\chi^i, 1]$, and is positive unless $I \to \infty$ and $\chi = 1$.

For a game $G$, now suppose that $\pi \leq (\beta + 1)/2\beta$. For any $\chi$ the probability of pluralistic ignorance can be made arbitrarily small for $I$ sufficiently large, and $0$ for $I \geq I_\pi$ with $I_\pi \geq 2$ increasing as $\beta$ decreases.

References


64


Assignment Of The State Of The World, An $I$-Tuple Of Types $(\theta_1, \ldots, \theta_I)$. Population $\psi$ and others’ types not observed; $\chi$ and $\pi$ common knowledge.
Figure II:

Timeline
Nature assigns the population: $\psi \in \{H, L\}$

Social planner $P$ privately observes $\psi$, makes public announcement $A \in \{H, L\}$

Nature assigns types $\theta_i \in \{H, L\}$ to individuals in the group. $P(\theta_i = \theta | \psi = \theta) > 1/2$

Player $i$ chooses action $a_i \in \{H, L\}$. Trades off choosing $\theta_i$ with how she is judged in period $i.5$

Others in the group silently judge player $i$: $J_{j,i} \in \{0, 1\}$ for all $j \neq i$

Figure III:
Timeline With Informed Policymaker