EXTREMISM AND SOCIAL LEARNING

Edward L. Glaeser[1] *and Cass R. Sunstein*[2]

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

ABSTRACT

When members of deliberating groups speak with one another, their predeliberation tendencies often become exacerbated as their views become more extreme. The resulting phenomenon — group polarization — has been observed in many settings, and it bears on the actions of juries, administrative tribunals, corporate boards, and other institutions. Polarization can result from rational Bayesian updating by group members, but in many contexts, this rational interpretation of polarization seems implausible. We argue that people are better seen as Credulous Bayesians, who insufficiently adjust for idiosyncratic features of particular environments and put excessive weight on the statements of others in situations of (1) common sources of information; (2) highly unrepresentative group membership; (3) statements that are made to obtain approval; and (4) statements that are designed to manipulate. Credulous Bayesianism can produce extremism and significant blunders — the folly of crowds. We discuss the implications of Credulous Bayesianism for law and politics, including media policy and cognitive diversity on administrative agencies and courts.

1. INTRODUCTION

Many people have celebrated the potential value of deliberation, including its uses in democracy (Habermas 1998), and it is tempting to think that group decision-making will both produce wiser decisions and average out individual extremism. In many settings and countries, however,

1 Fred and Eleanor Glimp Professor of Economics, Harvard University.

2 Felix Frankfurter Professor of Law, Harvard Law School. For valuable comments and suggestions, we are grateful to Daniel Benjamin, David Laibson, Eric Posner, Richard Posner, Andrei Shleifer, Adrian Vermeule, and participants in a workshop at Harvard Law School. We are also grateful to participants in a workshop at the National Bureau of Economic Research, above all our commentator Oliver Hart, and to an anonymous referee for an exceedingly careful report.

researchers have found that group deliberation leads people to take more extreme positions (Brown 1986). The increased extremism, often called *group polarization,* is usually accompanied by greater confidence and significantly decreased internal diversity, even when individual opinions are given anonymously (Schkade, Sunstein & Kahneman 2000; Brown 1986, 207). These facts, which are summarized in Section 2 of this article, appear to cast doubt on the wisdom, and certainly the moderation, of crowds. If deliberation leads liberals to become more liberal, and conservatives to become more conservative, the effects of deliberation are unlikely to be desirable in both cases. Deliberation might account for the folly, not the wisdom, of crowds.

2 Group polarization has evident implications for many issues in law and politics. It suggests, for example, that like-minded jurors, judges, and administrative officials will move to extremes. If group members on a corporate board or in a political campaign are inclined to engage in risk-taking behavior, group deliberation will produce increased enthusiasm for taking risks. But the mechanisms behind group polarization remain inadequately understood, and it is difficult to make predictions or to offer prescriptions without identifying those mechanisms.

3 In Section 3 of this article, we show that group polarization is predicted by a highly rational process of Bayesian inference. If individuals have independent information, which is shared in the deliberative process, then Bayesian learning predicts that ex post opinions will be both more homogeneous within the group and more extreme than individual opinions. Bayesian inference suggests that individuals with access only to their own private information will recognize their ignorance and hew towards the center. The information of the crowd provides new data, which should lead people to be more confident and more extreme in their views. Because group members are listening to one another, it is no puzzle that their post-deliberation opinions are more extreme than their pre-deliberation opinions. The phenomenon of group polarization, on its own, does not imply that crowds are anything but wise; if individual deliberators tend to believe that the earth is round rather than flat, nothing is amiss if deliberation leads them to be firmer and more confident in that belief.

4 While group polarization may reflect perfect Bayesian inference, there are other facts, summarized in Section 4, that cast doubt on this rosy ratio-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

nal interpretation. Often group deliberation produces greater confidence and greater extremism when little or nothing has been learned. Group polarization occurs even when little information is exchanged (Brown 1986). People appear to attend to the stated opinions of others even when those opinions are patently wrong (Asch 1955). Individuals often fail to give sufficient weight to the possibility that offered opinions are distorted by private incentives to mislead (Camerer, Ho & Chong 2004) or that people's actions reflect private information (Eyster & Rabin 2005). Outside the laboratory, professional persuaders, such as advertisers, political leaders, and clerics, have successfully led people to hold disparate religious beliefs that cannot all be true (Glaeser 2004), and to think, falsely, that they prefer the taste of Coke to Pepsi (Shapiro 2006) and that Mossad was responsible for the attacks of September 11, 2001 (Getzkow & Shapiro 2004).

In Sections 5–8 of this article, we suggest that social learning is often best characterized by what we call *Credulous Bayesianism*. Unlike perfect Bayesians, Credulous Bayesians treat offered opinions as unbiased and independent and fail to adjust for the information sources and incentives of the opinions that they hear. There are four problems here. First, Credulous Bayesians will not adequately correct for the common sources of their neighbors' opinions, even though common sources ensure that those opinions add little new information. Second, Credulous Bayesians will not adequately correct for the fact that their correspondents may not be a random sample of the population as a whole, even though a non-random sample may have significant biases.[3] Third, Credulous Bayesians will not adequately correct for any tendency that individuals might have to skew their statements towards an expected social norm, even though peer pressure might be affecting public statements of view. Fourth, Credulous Bayesians will not fully compensate for the incentives that will cause some speakers to mislead, even though some speakers will offer biased statements in order to persuade people to engage in action that promotes the speakers' interests. Our chief goal in Sections 5–8 is to show the nature and effects of these mistakes, which can make groups error-prone and anything but wise,

3 It is possible, of course, that a non-random sample will be unbiased; consider a non-random sample of neutral experts on the question whether, say, DDT imposes serious health risks. We use the term "representative sample" to mean relevantly representative, in the sense of lacking any kind of bias or skew.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

especially if they lack sufficient diversity.

6 In Section 5 of the article, we assume that errors in private signals are correlated across individuals. Credulous Bayesians overestimate the extent to which these signals are independent. The first proposition of the paper shows that when individuals are Credulous Bayesians, their post-deliberation beliefs become more erroneous and they acquire more misplaced confidence in those erroneous beliefs. This proposition helps explain why socially formed beliefs, like those about religion, politics, and constitutional law (and sometimes science as well), can be quite strongly held, despite a lack of evidence and an abundance of other groups holding opposing beliefs.

7 Our second proposition shows that when individuals are Credulous Bayesians, accuracy may decline as group size increases. As group size increases, mistakes can become more numerous and more serious. After all, the essence of Credulous Bayesianism is that people misuse the information of their neighbors, so more neighbors means more errors. This finding suggests that in some settings individuals may be wiser as well as less extreme than crowds (*compare* Surowiecki 2004; Page 2006).

8 We then turn to the possibility that an individual's friends and social networks are not a random sample of the population. A group of people might have skewed views on questions of policy or fact, and group members may not sufficiently adjust for that fact. We formalize this possibility by assuming that noise terms in the sample are correlated, rather than independent, as they could be if the group has been selected on some attribute or taste. Credulous Bayesians underestimate the correlation of the signals and act as if their neighbors are a random sample of the population as a whole. In this case, Credulous Bayesianism again causes more extremism and more error. Here too, larger group sizes (so long as they do not produce representativeness) can make decision-making less accurate. For a wide range of parameter values, more correlation decreases accuracy. This is our first result favoring intellectual diversity.

9 We then model intellectual diversity more formally as mixing people whose information reflects different group-specific error terms. Intellectual diversity is much more valuable for Credulous Bayesians than it is for perfect Bayesians. When a group moves from being totally homogenous to being formed out of two equally sized populations with different sources of common information, then the variance of the post-deliberation error falls

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

more quickly for Credulous Bayesians than it does for perfect Bayesians. Intellectual diversity always has value (Page 2006), but it becomes particularly important when decision-makers are Credulous Bayesians.

A large body of research has discussed the human tendency to give statements that conform to an expected community norm. For group deliberation, the problem is that people may discount this tendency and think, wrongly, that public statements actually convey information. In Section 6, we model conformism by assuming that individuals' statements reflect a combination of private information and an expectation of what individuals think that the group wants to hear. Credulous Bayesians fail fully to adjust for the fact that statements are skewed to the norm. The combination of conformism and Credulous Bayesianism creates error, tight homogeneity within groups, and greater heterogeneity across groups. If people utter politically correct statements, with the aim of avoiding the wrath of others, then Credulous Bayesianism could help explain both the blue state/red state phenomenon of ideological homogeneity within areas and heterogeneity across areas (Glaeser & Ward 2006).

In Section 7, we assume that some individuals, like legal advocates or politicians, have incentives to report misleading information in their quest to change people's decisions. "Polarization entrepreneurs," in law and politics, might attempt to do exactly that. This claim is in a similar spirit to Sendhil Mullainathan, Joshua Schwarzstein, and Andre Shleifer (2007), who examine the interaction between persuasion and categorical thinking. In this case, Credulous Bayesians fail fully to correct for the motives of those around them. The combination of incentive-created misstatements and Credulous Bayesianism always leads to less accurate assessment and can lead to bias as well. The degree of bias depends on the imbalance of resources or incentives across persuaders, not the persuasion per se.

Section 8 of the paper briefly considers some policy implications of Credulous Bayesianism. Of course it is true that identification of potential group errors cannot lead to any particular set of institutional arrangements, which must be chosen after considering many variables. But in the legal setting, our results help to explain the longstanding practice of requiring a degree of political diversity on the independent regulatory commissions, such as the National Labor Relations Board, and also cast light on the current debate over intellectual diversity on the federal judiciary.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

For public and private institutions, unbiased decision-making may depend more on maintaining balance across decision-makers than on the elimination of misleading statements. We also offer some brief remarks on media policy, with particular reference to the now-abandoned fairness doctrine. Our question here involves the likely consequences if people are exposed to beliefs from sources with a defined set of ideological convictions.

2. GROUP POLARIZATION: A GUIDED TOUR

13 The original psychological experiments on the effects of deliberation involved risk-taking behavior, with a demonstration that risk-inclined people become still more risk-inclined after they deliberate with one another (Stoner 1961). Risky decisions include taking a new job, investing in a foreign country, escaping from a prisoner-of-war camp, or running for political office (Hong 1978). With respect to many such decisions, members of deliberating groups became significantly more risk-inclined after a brief period of collective discussion. On the basis of this evidence, it became standard to believe that deliberation produced a systematic "risky shift" (Brown 1986).

14 Later studies drew this conclusion into serious question. On many of the same questions on which Americans displayed a risky shift, Taiwanese subjects showed a "cautious shift." Deliberation led citizens of Taiwan to become significantly less risk-inclined than they were before they started to talk (Hong 1978). Among American subjects, deliberation sometimes produced a cautious shift as well, as risk-averse people became more averse to certain risks after they talked with one another (Moscovici & Zavalloni 1969). The principal examples of cautious shifts involved the decision whether to marry and the decision whether to board a plane despite severe abdominal pain, possibly requiring medical attention. In these cases, the members of deliberating groups shifted not toward risk but toward greater caution (Moscovici & Zavalloni 1969).

15 A straightforward interpretation was able to reconcile these competing findings: the pre-deliberation median is the best predictor of the direction of the shift (*id.*; Brown 1986, 210–212). When group members are disposed toward risk-taking, a risky shift is observed. Where members are disposed toward caution, a cautious shift is observed. Thus, for example, the strik-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

ing difference between American and Taiwanese subjects is a product of a difference in the pre-deliberation medians of the different groups on the relevant questions (Hong 1978). Thus the risky shift and the cautious shift are both subsumed under the rubric of group polarization.

In the behavioral laboratory, group polarization has been shown in a remarkably wide range of contexts (Brown 1986; Turner *et al.* 1987, 142–170). Group deliberation produces more extreme judgments about the attractiveness of people shown in slides (Turner *et al.* 1987, 153). It also occurs for obscure factual questions, such as how far Sodom (on the Dead Sea) is below sea level (Turner *et al.* 1987). In a revealing finding at the intersection of cognitive and social psychology, groups have been found to make more, rather than fewer, conjunction errors (believing that A and B are more likely to be true than A alone) than individuals when individual error rates are high—though fewer when individual error rates are low (Kerr, MacCoun & Kramer 1996). Even burglars show a shift in the cautious direction when they discuss prospective criminal endeavors (Cromwell *et al.* 1991).

There is pervasive evidence of group polarization on issues that bear directly on politics and political behavior. With respect to affirmative action, civil unions for same-sex couples, and climate change, liberals become significantly more liberal as a result of discussion, while conservatives become significantly more conservative (Schkade, Sunstein & Hastie 2007). One experiment, conducted in Colorado, found that internal discussions among liberals in Boulder produce a strong shift to the left, resulting in both less internal diversity and more extremism; conservatives in Colorado Springs show a similar shift to the right (Schkade, Sunstein & Hastie 2007). In the same vein, white people who are not inclined to show racial prejudice show less prejudice after deliberation than before; but white people who are inclined to show such prejudice show more prejudice after deliberation (Myers & Bishop 1970). After deliberation, French people become more distrustful of the United States and its intentions with respect to foreign aid (Brown 1986, 224). Similarly, feminism becomes more attractive to women after internal discussions, at least if the relevant women are antecedently inclined to favor feminism (Myers 1975).[4]

4 There is a parallel literature that follows Lord, Ross & Lepper (1979) and that shows that initial views become more extreme after reading the same research. This finding — that people

16 Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

18 In the domain of law, there is considerable evidence of group polarization as well. Group polarization occurs for judgments of guilt and sentencing in criminal cases (Myers & Kaplan 1976; Kaplan 1977). In punitive damage cases, deliberating juries have been found to polarize, producing awards that are typically higher than those of the median juror before deliberation begins (Schkade, Sunstein & Kahneman 2000). When individual jurors begin with a high degree of moral outrage about a defendant's conduct, juries become more outraged, after deliberation, than their median member had been; but when jurors begin with little outrage about a defendant's conduct, juries become less outraged, after deliberation, than their median juror had been. Dollar awards are often as high as or even higher than the highest award favored, before deliberation, by any individual juror (Schkade, Sunstein & Kahneman 2000).

19 With respect to purely legal questions, panels of appellate judges polarize too. In ideologically contested areas (involving, for example, disability discrimination, sex discrimination, affirmative action, environmental protection, and campaign finance regulation), Republican appointees show especially conservative voting patterns when sitting on panels consisting entirely of Republican appointees, and Democratic appointees show especially liberal voting patterns when sitting solely with other Democratic appointees (Sunstein *et al.* 2006; Sunstein, Schkade & Ellman 2004). This pattern has proved highly robust; it has been found in over twenty areas of substantive law. As a result, panels consisting of three Republican appointees show radically different voting patterns from those of three Democratic appointees, in a way that seems to ensure that political judgments compromise ideals associated with the rule of law (Sunstein, Schkade & Ellman 2004). In short, federal judges show the same polarization effects as do liberals and conservatives in Colorado.

3. GROUP POLARIZATION AND BAYESIAN INFERENCE

20 Why does group polarization occur? In this section, we show that the preced-

with different beliefs disagree more after being exposed to the same information — is often said to involve "biased assimilation" of new material; it is harder to reconcile with standard Bayesian learning than the finding that group discussion increases polarization.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

ing facts about group polarization are not only compatible with Bayesian inference, but are primary predictions of the standard Bayesian model of social learning. We assume that individuals are trying to form an assessment of an unknown parameter, *D,* which might reflect the damages in a civil trial, the proper resolution of a constitutional dispute, the right response to climate change, or the dishonesty of a political candidate. For some readers, the exposition will seem somewhat technical, but the intuition should not be obscure: If group members are listening to and learning from one another, their discussions will produce greater confidence. As confidence increases, convictions become firmer and individuals are more comfortable moving away from the center in the same direction as their predeliberation tendencies.

Throughout this paper, we will make use of the normal signal extraction formula and therefore assume that all random variables are normally distributed. The true value of D has mean zero (an arbitrary normalization) and variance $\frac{1}{P_0}$. Each individual receives a private signal, denoted S_i for individual i, which equals D plus a noise term, denoted η_i, which is also mean zero and has variance $\frac{1}{P_1}$. Bayes' rule tells us that if an individual had nothing but this private signal, that individual's estimate of D would equal $\frac{P_1 S_i}{P_0 + P_1}$.

Our interest lies in social settings where "I" individuals communicate a share of their private signals. In this first setting, we assume that people just relay their signals accurately to the group and that all of the signals are independent. The signal extraction formula then tells us that after communication all I individuals will share the same posterior assessment of D, which equals

$$\frac{P_1 \sum_i S_i}{P_0 + IP_1} \text{ or } \frac{IP_1 D + P_1 \sum_i \eta_i}{P_0 + IP_1}$$

(*see, e.g.,* Thomas Sargent 1979). This equals D plus an error term,

$$\frac{P_1 \sum_i \eta_i - P_0 D}{P_0 + IP_1}.$$

One of the most empirically problematic implications of the Bayesian model is that people will end up with the same beliefs, even if they start with different signals. Still the Bayesian model delivers two standard facts about the heterogeneity of posterior beliefs:

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Claim #1: As the number of people that communicate independent signals increases, the posterior will become more accurate, in the sense that the variance of the error will fall, and the variance of posterior beliefs will rise.

24 The variance of error term equals $\frac{1}{P_0 + IP_1}$ which is obviously declining with I. The variance of the posterior equals $\frac{IP_1}{P_0(P_0 + IP_1)}$, which is increasing with I. The ex post evaluation equals $\frac{IP_0 + IP_1}{P_0 + IP_1}$ times the average evaluation ex ante, which is obviously greater than one. Group assessments are a simple multiple of average individual assessments; it follows that just as in the experiments discussed above, we should expect to see groups head towards extremes in the direction suggested by individual opinions.

25 Moreover, it is quite easy to think of examples, especially when I is small, where the ex post opinion of the group is more extreme than even the most extreme antecedent opinion among the group's members. Assume that there were three people with private signals of 0, 1, and 2 and assume that $P_0 = 4$ and $P_1 = 1$. Before the information exchange, the most extreme individual assessment of D was 0.4, which equals two times $\frac{P_1}{P_0 + P_1}$ or 0.2. After the information exchange, everyone believes that the expected value of D is 0.429, which equals one (the average signal) times $\frac{IP_1}{P_0 + IP_1}$ or $\frac{3}{7}$. The group is not only more extreme than the average individual opinion but also more extreme than the most extreme individual opinion.

26 These findings rationalize the results of David Schkade, Cass Sunstein & Reid Hastie (2007), and show that information exchange should lead to more extremism and less internal diversity, even when people are perfectly rational. Recall that when residents of Colorado Springs, a relatively conservative place, are brought together for group discussion, their opinions become more conservative on climate change policy, affirmative action, and civil unions for same-sex couples. When residents of Boulder, a relatively liberal place, come together, their opinions become more liberal on all three issues. In the context of the model, this experiment is akin to selecting a group that has positive signals and a group that has negative signals. If the two groups believe that they have been randomly formed, and not selected on the basis of their signals, then both groups will become more extreme. This type of extremism from the exchange of information and opinions is exactly what a rational model with group learning would predict.

27 The standard explanations of group polarization show an intuitive

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

appreciation of some of the more formal analysis here. One of the most prominent explanations refers to "informational influences" (Brown 1986). This account is essentially Bayesian. The central idea is that group members will be aware of some, but not all, of the arguments that support the antecedent tendency within the group. As information is pooled, learning occurs, and the learning will predictably tend to intensify the antecedent tendency. It is evident that polarization often occurs as a result of such learning.

Other explanations of group polarization, invoking the effects of confidence and corroboration on people's preexisting views, are also quite Bayesian. If rational people lack confidence, they will tend toward the middle and hence avoid the extremes. As people gain confidence after hearing their views corroborated by others, they usually become more extreme in their beliefs (Baron *et al.* 1996, 557–559). This is exactly what the Bayesian model suggests. In a wide variety of experimental contexts, people's opinions have been shown to become more extreme simply because their views have been corroborated, and because they have become more confident after learning that others share their views (Baron *et al.* 1996). The Bayesian model predicts this process.

If group polarization merely reflects standard Bayesian inference, then it is hard to think that there is anything about polarization that challenges the conventional view that groups make better decisions than individuals. If people become firmer in their conviction that cigarettes cause cancer, that it is risky to drink and drive, or that the earth rotates around the sun, nothing is amiss. In the next section, we suggest that there are facts that are less compatible with rational Bayesian inference and more compatible with less perfect social learning. We will then suggest that a more realistic model of social learning presents a somewhat different and far less positive view about the accuracy of group decision-making. In this way, we can obtain a better understanding of the occasional folly of crowds.

4. CREDULOUS BAYESIANS

The most obvious problem with the Bayesian framework is that group polarization is found even when people learn essentially nothing. To be sure, people must learn the bare fact that other people hold certain positions,

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

and one can learn something from that fact, so long as others are not unreliable. But a significant shift in the domains of law and politics — toward, for example, enthusiasm for affirmative action policies or an international agreement to control greenhouse gases — should not be expected when people have discussions with others who lack independent information.

31 Return once more to the Colorado experiment. Both liberals and conservatives regularly polarized on questions involving climate change, affirmative action, and civil unions for same-sex couples, even though much of the time the discussion produced little or no new information. The same is true in the standard risky-shift experiments: People become more risk-inclined simply after learning that others are risk-inclined, without learning much, if anything, about why it makes sense to take risks. Indeed, group polarization occurs when people are merely exposed to the conclusions of others, rather than to the reasons for those conclusions (Brown 1986).

32 In many of the discussions in Colorado Springs and Boulder, the individuals who talked to each other were not bringing new data to the table. The residents of both of these places had been exposed to the same basic influences throughout most of their lives. For at least two of these issues (affirmative action and same-sex unions), there has been a huge amount of public discourse, and the discussions tended to add essentially nothing. For one of them (climate change), the issues are technical and complex, and few participants were actually able to provide new information. True, we have emphasized that people learn from others on technical and nontechnical issues alike; most people begin with little memory of relevant facts and something is conveyed by the mere opinion of another human being who seems reliable. Bare conclusions can themselves be informative, especially but not only if they come from experts. But it seems heroic to suggest that the dramatic movements we observe are solely or even mostly a product of new knowledge. So while the move to extremes is exactly what one might have predicted with a Bayesian model of social learning, the actual situation suggests a somewhat different scenario: people are treating each other's opinions as offering new information even when there actually is nothing new.

33 Similar problems occur with models of group polarization that do not emphasize acquisition of information through discussion. According to the "social comparison" account, offered as an alternative to accounts that

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

emphasize information, people want to maintain both their preferred self-conception and their preferred self-presentation (Brown 1986). Suppose, for example, that group members believe that they are somewhat more likely to take investment risks than other people, or that they are somewhat more critical of the incumbent president than other people. If such people find themselves in a group of people who inclined to take investment risks, or to criticize the incumbent president, they might shift, not because they have learned anything, but because they want to see themselves, and present themselves, in the preferred way. The social comparison account makes a place for conformism, taken up in detail below. What that account misses is that in many settings, people shift in the belief that others are saying what they think or know, even though they are actually motivated by social motivations. Sometimes people do move because they desire to conform, but sometimes they move because they believe, falsely, that the views of others provide valuable information.

In the remainder of the article, we explore an alternative explanation of group polarization: the hypothesis that individuals are Credulous Bayesians, who treat proffered opinions as having significant information value, even when those opinions are biased or based on no new information. Of course most heuristics work reasonably well (Tversky & Kahneman 1971; Gigerenzer 2007), and there is nothing inherently unreasonable about Credulous Bayesianism, which is essentially a Bayesian heuristic. Most of the opinions that we hear over the course of a day are given to us by well-meaning people who are in fact sharing new information, on topics like the quality of a restaurant or whether or not it is raining. It is quite sensible to take those opinions seriously. It is even more sensible for children to be attuned to take the advice given to them by their parents, on, say, not playing with knives or not eating poisonous food. In fact, the human ability to learn from one another is a cornerstone of our species and its civilizations.

With this background, we can see that Credulous Bayesianism is essentially a failure to fine-tune learning for each individual setting, and a failure to correct for the motives of others. We may, on average, assign the right amount of weight to the opinions of friends or strangers, but in some settings we should be more skeptical, but do not bother to think through the appropriate degree of skepticism. A sensible model of costly cognition

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

could easily explain such a tendency, especially in settings where getting it right has little private value. It may well be that the tendency to be suspicious imposes costs beyond mere thinking, because suspicion may also be linked to uncomfortable emotions and an inability to enjoy a pleasant conversation. We suspect that going through life being intensely suspicious of every new statement would impose enormous cognitive and emotional costs. Credulous Bayesians are able to avoid such costs.

36 Credulous Bayesianism is linked to the cognitive hierarchy model of Colin Camerer, Teck-Hua Ho & Kuan Chong (2004). Camerer, Ho & Chong show that behavior in many games can be rationalized with a model in which only a subset of the population is able to think about the motives and implied actions of those around them. Credulous Bayesianism is quite similar in that when people use it, they are essentially failing to think through why opinions that they hear may not be unbiased or informative. There is also a link between Credulous Bayesianism and the Cursed Equilibrium concept of Erik Eyster and Matthew Rabin (2005). In a Cursed Equilibrium, individuals make a naïve inference based on people's actions; Credulous Bayesians make a naïve inference based on people's statements.

37 We believe that Credulous Bayesianism makes more sense of the Colorado experiment than perfect Bayesian inference. Little new information was proffered in this setting, so it is hard to see how Bayesian learning occurred, but opinions were given. If Credulous Bayesianism makes us prone to put weight on the opinions of others (even when they are based on the same sources as our own views), then we would expect to see something that looks like learning, in that opinions become more extreme and more tightly held, even though little or no real information is being exchanged.

38 In fact, an abundance of psychological evidence suggests that people do use a "follow the majority" heuristic that puts great weight on the opinions of others (Gigerenzer 2007). Psychologists have often shown that people follow the views of others even when those others are palpably wrong. In the most famous experiments, Solomon Asch explored whether people would be willing to overlook the apparently unambiguous evidence of their own senses (Asch 1955). In these experiments, the subject was placed into a group of seven to nine people who seemed to be other subjects in the experiment but who were actually Asch's confederates. The simple task was

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

to "match" a particular line, shown on a large white card, to the one of three "comparison lines" that was identical to it in length. Asch's striking finding was that when confronted with the obviously wrong but unanimously held views of others, most people end up yielding to the group at least once in a series of trials. Indeed, in a series of twelve questions, no less than 70 percent of people went along with the group, and defied the evidence of their own senses, at least once (Asch 1955).

It might seem jarring to think that people would yield to a unanimous group when the question involves a moral, political, or legal issue on which they have great confidence. But additional experiments, growing out of Asch's basic method, find huge conformity effects for many judgments about morality, politics, and law (*see* R.S. Crutchfield 1955). Such effects were demonstrated for issues involving civil liberties, ethics, and crime and punishment. Consider the following statement: "Free speech being a privilege rather than a right, it is proper for a society to suspend free speech when it feels threatened." Asked this question individually, only 19 percent of the control group agreed. But confronted with the shared opinion of four others, 58 percent of people agreed (Krech, Crutchfield & Ballachey 1962, 509).

Non-experimental evidence also supports the view that people put weight on opinions that they hear, even when those opinions are biased by incentives or based on very limited information. Billions of dollars are spent on advertising, only some of which is informative, and much of that advertising contains obviously biased statements that are made about a product's qualities. If these statements were not effective, it would be hard to believe that firms would spend so much on them. There is also more direct evidence on the impact of advertising. People generally say that they think they prefer the taste of Coke to Pepsi, and their buying habits suggest that they believe that to be true, yet in blind taste tests people generally prefer the taste of Pepsi to Coke (Shapiro 2006).

Other settings also seem to suggest that people are Credulous Bayesians. There is a paucity of hard information about many religious topics, including the afterlife, the nature of any deities that may exist, and the lives or even the existence of many historical religious figures. Rational Bayesians should have very weak beliefs about things on which so little tangible evidence ex-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

ists. Yet individuals are strongly committed to their religious beliefs and are often willing to die or kill for them. These beliefs are greatly influenced by statements of parents or religious leaders, and both of these groups often have little direct knowledge of core religious beliefs, but strong incentives to shape those beliefs in a particular direction. A more perfect Bayesian would presumably treat these opinions with a great deal of skepticism, especially when that Bayesian learns that millions of others hold very different beliefs. The strong beliefs about religion that are transferred from person to person look to us more like Credulous Bayesianism.

42 In political settings, there is also evidence that people accept statements that they hear, with relatively little critical appraisal. Matthew Getzkow and Jesse Shapiro (2004), for example, document the remarkable range of beliefs held in the Islamic world about Israel and the United States. Most striking among those is the view that Mossad was responsible for the 9/11 terrorist attacks on America. Within the United States as well, there is abundant evidence that many people have acquired incorrect beliefs, such as a vastly overinflated view of the percentage of the national budget that the U.S. government spends on foreign aid or the connection between Saddam Hussein and the 9/11 attacks.[5] Again, the heterogeneity in these views seems incompatible with any kind of perfect Bayesianism, but it is much easier to reconcile with individuals who are Credulous Bayesians.

5. CREDULOUS BAYESIANS, COMMON NOISE, AND BIASED SAMPLES

43 We now turn to our theoretical exploration of Credulous Bayesianism. To reflect the core rationality of social learning, we begin with the Bayesian learning model described in Section 3, where individuals are trying to infer the truth based on a combination of their private information and the information sent by others. The critical deviation from standard Bayesian learning models is that we assume that *people do not accurately assess the degree to which social signals actually represent new information*. We are essen-

5 An alternative interpretation of Americans' routine overestimation of the percentage of spending on foreign aid is that people routinely overestimate the percentage of spending on small budget items.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

tially following Peter DeMarzo, Dimitri Vayanos and Jeffrey Zwiebel (2003) by assuming that individuals have trouble correcting the degree to which they should down-weight social signals in any given setting.

In this first model, we follow DeMarzo, Vayanos and Zwiebel (2003) closely and consider the case where individual opinions share a strong common component, but the degree of common noise is underestimated by the participants in the model. As in Section 3, we think of a group of I individuals sharing information and forming a posterior assessment of the value of D. This setting could capture the deliberations of a judicial panel or any group of people that exchanges ideas and then makes a decision.

As in Section 3, we assume that individuals receive a private signal equal to D plus a noise term denoted η_i. However, unlike in Section 3, we now assume that the noise term η_i is the sum of a common noise term θ and an individual specific term ε_i. The total variance of the noise term remains $\frac{1}{P_1}$. The variance of θ is $\frac{\nu}{P_1}$ and the variance of ε_i is $\frac{1-\nu}{P_1}$. This change does not alter the Bayesian formula for pre-deliberation opinions, which remains $\frac{P_1 S_i}{P_0+P_1}$.

If individuals shared their signals and used the correct Bayesian updating formula, their posterior view would equal

$$\frac{P_1 \sum_i S_i}{P_0(1+(I-1)\nu)+IP_1}.$$

This formula puts less weight on the signals of others as the share of the noise that is common increases. Since the share of common noise makes each new signal less informative, people appropriately put less weight on each new signal when much noise is common.

Throughout the remainder of this paper, we assume one common core deviation from perfect Bayesianism: individuals underestimate the degree to which information that they hear from others is biased or based on common sources. In this context, we model this underestimation by assuming the people underestimate the extent to which the noise terms are common. Specifically, every one of the agents in the model believes that the variance of θ is $\frac{\lambda\nu}{P_1}$ and the variance of ε_i is $\frac{1-\lambda\nu}{P_1}$. The parameter λ measures the extent to which individuals are Credulous Bayesians and underestimate the amount of common noise. The parameter ranges from one, which would mean perfect rationality, to zero, which would mean a complete failure to

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

assume any common error terms. We also assume that individuals update as if they knew this parameter perfectly, so that the new formula for posterior beliefs is

$$\frac{P_1\sum_i S_i}{P_0(1+(I-1)\lambda\nu)+IP_1} \text{ or } \frac{P_1(\sum_i \varepsilon_i + I(D+\theta))}{P_0(1+(I-1)\lambda\nu)+IP_1}.$$

48 We first discuss a simple example where people are completely credulous, that is, λ is zero. The average belief of I people who have not shared their information will equal $\frac{P_1\sum_i S_i}{I(P_0+P_1)}$. The average belief after signals are shared will equal

$$1+\frac{(I-1)(l-\lambda\nu)P_0}{P_0(1+(I-1)\lambda\nu)+IP_1}.$$

Whatever initial opinions that existed in the group will get more extreme as the group size expands. If $D=0$ then groups that initially erred in thinking that D is positive will become more extreme in their belief and groups that initially erred in thinking that D is negative will get more extreme in that belief.

49 The experiments in which conservatives and liberals are grouped together can be seen as an example of that phenomenon. In this case, the experimenter has structured the two groups so that one group's signals are positive and the other group's signals are negative. The model predicts that both groups will get more extreme. The liberals will get more liberal and the conservatives will get more conservative. Of course, to use this model at all, we must assume that individuals are unaware that their group has been chosen on the basis of its signals.

50 This result holds for sophisticated Bayesians as well. However, Credulous Bayesianism increases the tendency of priors to become extreme. In either case, if an individual is thrown in with a random group, the expected impact of that group on the individual's bias is zero. However, if the group starts with an average bias that is positive or negative, then communication across that group will make this bias more extreme. We are particularly interested in the impact of Credulous Bayesianism on three different variance terms associated with ex post beliefs. First, we are interested in the actual variance of ex post beliefs, which equals

$$\frac{IP_1(P_0(1+(I-1)\nu)+IP_1)}{P_0(P_0(1+(I-1)\lambda\nu)+IP_1)^2}.$$

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Higher levels of variance of posterior beliefs mean more polarization of beliefs. Second, we are interested in the variance of the true error associated with those beliefs or

$$\frac{(1-\lambda)(I-1)\nu P_1 I}{(P_0(1+(I-1)\lambda\nu)+IP_1)^2}+\frac{1+(I-1)\lambda\nu}{P_0(1+(I-1)\lambda\nu)+IP_1}.$$

The variance of this error term captures the extent to which people are incorrect in the posterior assessments. Finally, we are interested in the variance of the error term that people believe to be associated with their posterior beliefs, or

$$\frac{1+(I-1)\lambda\nu}{P_0(1+(I-1)\lambda\nu)+IP_1}.$$

This error term reflects the degree of certainty that people have ex post.

The following proposition gives the impact of Credulous Bayesianism on variance in posterior beliefs, on the variance of the error associated with those beliefs, and on the perceived variance of the error associated with those beliefs:

> *Proposition 1:*
>
> As the value of λ falls, the variance of ex post beliefs increases, the variance of the error in those beliefs also rises, and the degree to which people underestimate their true level of error rises as well. As the share of the noise term that is common increases, the difference between the actual precision of the posterior and the perceived precision of the posterior will rise as long as $P_0+IP_1 > (I-1)\lambda\nu P_0$.

Proposition 1 shows that Credulous Bayesianism causes the heterogeneity of posterior beliefs to rise, which means that group polarization will be more pronounced among Credulous Bayesians. Belief in the independence of information makes it more likely that groups will converge on a more extreme belief, essentially because they believe that they have better information than they do. In fact, the evidence is consistent with this finding (Brown 1986; Schkade, Sunstein & Kahneman 2000).

While the extremism associated with standard Bayesian learning accompanies an increase in accuracy, Credulous Bayesianism increases extremism while decreasing accuracy. The fact, stated in Proposition 1, that Credulous Bayesianism causes the variance of the error term to increase means that on

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

average posterior beliefs will be less accurate. This is not surprising; excessive credulity is, after all, going to produce errors. It is perhaps somewhat more surprising that the perceived variance of the posterior error falls with Credulous Bayesianism. This means that Credulous Bayesianism can explain the phenomenon of people holding onto erroneous opinions with a great deal of intensity.

54 The final sentence in the proposition shows that the degree of erroneous confidence may rise as the noise terms become more highly correlated across people. This effect is not automatic because increasing the amount of common noise creates two opposing effects. First, the increase in common noise does cause individuals to put less weight on social signals. Second, as the amount of common noise rises, the tendency to think that each person has independent information becomes costlier. If λ is sufficiently low, so that the bias is sufficiently severe, then the second effect must dominate, and more common noise means more false confidence. This fact suggests that Credulous Bayesians will make particularly bad decisions when there is common noise.

55 Proposition 1 tells us that decision-making gets worse as people mistakenly fail to correct for the common sources of information, but it does not show that group decision-making can actually be worse than individuals acting alone. Proposition 2 gives conditions in which the level of error, as measured by the variance of the error of the posterior, can rise with group size:

> *Proposition 2:*
>
> The variance of the error of the posterior estimate will always decline with I if λ is sufficiently close to one. If $I > \frac{P_0 + P_0\nu}{2\nu P_0 - P_1(1-\nu P_1)}$ and λ is sufficiently close to zero, then the variance of the error of posterior beliefs will rise with I.

If people are sufficiently prone to Credulous Bayesianism, that is, if λ is sufficiently close to zero, and if common noise is sufficiently important, then group size decreases accuracy. These conditions do not imply that the folly of groups is a general condition. In many cases these conditions will fail, and groups will be wiser than individuals, but the possibility that the conditions of Proposition 2 will sometimes be met should warn us that groups will sometimes make worse decisions than individuals. And in fact, a great deal of evidence suggests that groups often perform worse than their best member or even their median member (Gigone & Hastie 1995).

56 We acknowledge a glaring failure of the Credulous Bayesian model as

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

it now stands, which is that it predicts that when two disparate groups are brought together, their views will converge. Certainly, such convergence can occur, but it is not the general rule. In many cases, members of a deliberating group will shift toward a more extreme direction in line with the predeliberation median, even if those members come from two disparate groups (Sunstein 2003). At least if people's views are not entrenched, a group consisting of three people who tend to think that climate change is a serious problem, and two who tend to reject that view, will often shift toward more concern about climate change (a prediction consistent with group polarization). But there also are abundant examples of cases in which group members stick to their extreme opinions when they connect with each other (Brown 1995; Sunstein 2003). In other words, a group that consists of equally opposed subgroups might well show neither convergence nor polarization but simply entrenchment of members' antecedently held views.

For example, connections between different religions rarely leads to a merging of religious beliefs. And if (certain) Palestinians meet with (certain Israelis), convergence is not expected. Entrenchment and continuing conflict are at least as likely. We conjecture that the model could generate permanent disagreement if people put a high weight on the views of "insiders" but believe that the opinions of "outsiders" are essentially worthless. If there is sufficiently little trust across groups, then it may be impossible to communicate information effectively. In that case, insiders may share views and extremes may persist even after communication (Brown 1986). If insiders in each subgroup listen to one another, and discount the views of those in the other group, then members of each group will polarize, producing an even greater split between the two.

5.1. Biased Samples

To illustrate the impact of biased samples of correspondents, we now assume that individual signals include a common and an idiosyncratic component. This represents a special case of the correlated errors discussed above, but it is a case that makes it particularly easy to discuss sample selection. The correlation of error terms within a group would occur if the group were not a random sample of the world, but rather a specific subset with highly similar backgrounds or biases. If people sorted into communities or conversed with people who had similar views, then we should certainly expect to see

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

correlation in the error terms of people who regularly communicate with one another. The intuition here is straightforward: People who find themselves in a certain group often do not give sufficient thought to the possibility that the group is unrepresentative in a way that ought to matter for purposes of social learning. People in left-leaning or right-leaning groups listen to fellow group members, without discounting what they learn in light of the dispositions of those members.

59 Perfect Bayesians should strongly discount the opinions of their highly selected acquaintances. Even if they are fairly committed to their group identity, rational people should know that there is a skew in the information held within group members (Democrats, Republicans, union members, Americans, and so forth). Credulous Bayesians, however, imperfectly correct for the unrepresentative nature of the opinions of their friends and acquaintances. They assume that their friends and acquaintances are representative of the world, not a very unusual set of informants.

60 To address this possibility, we assume that individuals continue to receive a private signal equal to D plus a noise term denoted η_i. The variances of these two terms are the same as before, but now we assume that the covariance of any two noise terms in a person's group is $\frac{\psi}{P_1}$. With this assumption and assuming that people know this covariance term, the correct Bayesian inference formula is

$$\frac{P_1 \sum_i S_i}{P_0(1+(I-1)\psi)+IP_1}.$$

As the degree of correlation rises, individuals should put less weight on the opinions of their neighbors.

61 In this instance, Credulous Bayesians underestimate the degree to which their neighbors' error terms are correlated with their own, which is equivalent to assuming that one's neighbors are more representative of the overall population than they actually are. To capture this, we assume that people believe that the true covariance of neighbor's signals is $\frac{\lambda\psi}{P_1}$. Credulous Bayesians use an updating formula of

$$\frac{P_1 \sum_i S_i}{P_0(1+(I-1)\lambda\psi)+IP_1}$$

which puts too much weight on the signals received by a biased sample of neighbors.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

The impact of Credulous Bayesianism when there are correlated noise terms is almost identical to the impact of Credulous Bayesianism when there is a common noise term.

Proposition 3:

(a) As the value of λ falls, the variance of ex post beliefs increases and the variance of the error in those beliefs also rises.

(b) The variance of the posteriors will always rise with I.

(c) The variance of the error of the posterior estimate will decline with I as long as λ is close enough to one. If

$$I > \frac{(1+\psi)P_0}{2\psi P_0 - (1-\psi)P_1},$$

then the variance of the error of the posterior estimate will rise with I as long as λ is close enough to zero.

Just as before, individuals who fail to understand the degree to which their group is not representative will tend to have more extreme and more erroneous posterior beliefs. Credulous Bayesians have excessive faith in the views of their neighbors and so underestimate the extent to which their neighbors make mistakes that are similar to their own. The folly of crowds is a likely result.

The third part of the proposition shows that when individuals are close to being perfect Bayesians, then a larger group will always lead to views that are more accurate, as well as more extreme. However, if individuals are close to being completely naïve Bayesians, who don't think that there is any correlation in their signals, then as long as

$$(2-\frac{1}{I})\frac{\psi}{1-\psi} - \frac{1}{I(1-\psi)} > \frac{P_1}{P_0},$$

accuracy will decline with group size. These are sufficient conditions for crowds to be more foolish than individuals. The intuition behind the condition is that this somewhat perverse result is more likely when signals are highly correlated, when group size is already large, and when the signal to noise ratio in individual signals is quite low.

The next proposition provides our first result on the benefits of diversity, which in this case means a reduction in the degree of correlation across signals.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proposition 4:

(a) As the degree of correlation in the signal noise (ψ) rises, the variance of the posterior belief will decline if and only if

$$\lambda > \frac{IP_1 + IP_0}{2IP_1 + (2 + \psi(I-1)P_0}.$$

(b) The variance of the error in the posterior belief will rise with ψ if and only if

$$\frac{IP_1 + P_0}{\psi(I-1)P_0} > \lambda(1-2\lambda).$$

66 The first part of this proposition shows that increasing correlation of signals can either increase or decrease ex post extremism. Part (a) implies that when λ is greater than one half, then increasing correlation of signals will cause posterior beliefs to become less extreme. There is nothing surprising in this fact. Lower correlation of noise means that there is more information in the signals, and more information should cause beliefs to become more extreme.

67 The opposite result occurs when λ is sufficiently less than one half. In that case, more correlation actually leads to more ex post extremism. High correlation of noise means that the average signal will get more extreme, and since people are putting too much weight on that average signal, their views get more extreme as well.

68 The second part of the proposition gives a condition under which a reduction in correlation in the noise, which can be seen as more diversity, increases the accuracy of ex post beliefs. If λ is greater than one half or sufficiently low, then this natural result always holds. More diversity improves accuracy. The opposite result can occur when people reason particularly poorly, but we suspect that this is more of a theoretical curiosity than a case to worry about.

5.2. Intellectual Diversity

69 We now continue with our investigation of intellectual diversity and return to the framework with common noise terms. This simple framework is meant to help evaluate the benefit of intellectual diversity in groups, such as panels of judges, legislatures, or students in a classroom. Our specific interest is to ask whether diversity is more or less valuable as people become more rational.

70 We model diversity by assuming that different groups have independent draws of the common noise term θ. While such diversity always improves

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

information quality, the value of diversity can be significantly higher when people are Credulous Bayesians. We assume that a fraction α of the I people who are sharing information come from a group that has one draw of θ while the rest of the group has received a second independent common signal. The true Bayesian posterior is then:

$$(1)\ \frac{P_1(\sum_{i<\alpha I}(1-\nu+\nu I(1-\alpha))S_i+\sum_{i>\alpha I}(1-\nu+\nu I\alpha)S_i)}{((1-\nu)^2+(1-\nu)\nu I+\alpha(1-\alpha)\nu^2 I^2)P_0+((1-\nu)I+2\alpha(1-\alpha)\nu I^2)P_1}.$$

Credulous Bayesians recognize that people from the two groups have different common noise terms, but they continue to underestimate the true amount of common error by a factor λ and this produces the modified formula:

$$(1')\ \frac{P_1(\sum_{i<\alpha I}(1-\lambda\nu+\lambda\nu I(1-\alpha))S_i+\sum_{i>\alpha I}(1-\lambda\nu+\lambda\nu I\alpha)S_i)}{((1-\lambda\nu)^2+(1-\lambda\nu)\lambda\nu I+\alpha(1-\alpha)\lambda^2\nu^2 I^2)P_0+((1-\lambda\nu)I+2\alpha(1-\alpha)\lambda\nu I^2)P_1}.$$

This formula nests both the perfect Bayesian and a learner who completely ignores the common error components in people's signals. This ultra-naïve Bayesian would use the formula for learning with independent signals, $\frac{P_1\sum_i S_i}{P_0+IP_1}$. As long as $\lambda > 0$, then Credulous Bayesians, like perfect Bayesians, put more weight on the views of the members of the minority group since that group provides more information as its members had a different common shock. In Section 9, we will turn to a case in which individuals put less weight on the opinions of outsiders.

The next proposition discusses the impact of increasing intellectual diversity by changing the population shares of the two groups:

Proposition 5:

(a) The variance of prediction error is declining with α if and only if $\alpha < 0.5$ for both perfect Bayesians and naïve Bayesians. More generally, the variance of prediction error is declining with α when α is small enough and the other parameters satisfy

$$0 > IP_1(-2+\lambda\nu(4-2\lambda\nu+I(-2+\lambda+\lambda\nu)))+$$
$$P_0(-2+\lambda\nu(2+\lambda(-3I(1+\nu)+2(2+\nu)-(I-1)\nu(-8+I+(I-2)\nu)\lambda+2(I-2)$$
$$(I-1)\nu^2\lambda^2))).$$

(b) An increase in α from 0 to 0.5 will cause a decline in the variance of the error of posterior beliefs for naïve Bayesians that is larger than the decline in the variance of the error for perfect Bayesians.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889922 by guest on 20 March 2024

Part (a) of the Proposition tells us that when α is small, an increase in the amount of diversity increases the accuracy of posterior beliefs when people are either perfect Bayesians or naïve Bayesians. A greater mix of people makes it easier to factor out the common noise for either of these two extreme groups. The condition given in the proposition is sufficient to ensure that diversity will be good in more intermediate cases, and it seems likely to hold for most reasonable parameter values.

72 The second part of the proposition tells us that the advantages associated with mixing will be larger when people are Credulous Bayesians than when people are perfect Bayesians. When there is a lot of common noise, the naïve Bayesians suffer both because of that noise and because they misattribute that noise to underlying truth. As there is more mixing, the common noise gets averaged out, and there is both more accuracy and less misattribution among the naïve Bayesians. This result suggests that intellectual diversity is particularly valuable when people incorrectly underestimate the common source of signals. Of course the benefits of diversity would fall if members of one group did not trust the statements of members of another group; in that event, those statements might be unhelpful or even counterproductive.

6. CREDULOUS BAYESIANS AND CONFORMISM

73 At this point, we drop the assumption that individuals perfectly report their own signals and instead assume that individuals exhibit a tendency towards conformism when reporting their results. We also assume that all errors are idiosyncratic. This tendency towards conformism might come about for signaling reasons, as in Stephen Morris (2001), or out of a taste for conformity, as in B. Douglas Bernheim (1994). We assume that people care both about reporting the truth and about saying something that conforms to the norm in their group. The problem arises when people do not sufficiently discount people's statements, treating those statements as informative when in fact they reflect only the pressure to conform. In a political organization, for example, group members may disregard the possibility that disparagement of some environmental concern is driven not by knowledge, but by a perception that disparagement of environmental concerns is popular within the relevant group.

74 It is true of course that people can introspect and realize that they are

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

themselves conformists; they might generalize from their own self-knowledge to discount the behavior of others as conformists too. But evidence of human conformity constantly produces real surprise, and in any case human beings are subject to "the fundamental attribution error" (Ross & Nisbett 1991), which means that they tend to attribute people's behavior to their dispositions, rather than to the situation (while simultaneously attributing their own behavior, at least when it is good, to the situation, rather than their dispositions). It follows from the fundamental attribution error that people are unlikely to see the extent to which other people's behavior is a product of the desire to conform. Conformity pressures are responsible for "pluralistic ignorance," understood to mean ignorance of the judgments and beliefs of other people. The group norm distorts what people say they believe.

The norm is known and denoted η. Specifically, we assume that people report a signal meant to minimize the quadratic loss functions: $\gamma(\tilde{S}_i - S_i)^2 + (1-\gamma)(\tilde{S}_i - \eta)^2$ that sums losses from lying and losses from deviating from the community norm. Optimizing behavior then generates the reporting rule: $\tilde{S}_i = \gamma S_i + (1-\gamma)\eta$, which means that reported signals are an average of the truth and the community norm. If people correctly transform reported signals, by subtracting $(1-\gamma)\eta$ and multiplying by $\frac{1}{\gamma}$, then use of the standard Bayesian inference formula

$$\frac{P_1(S_i + \frac{1}{\gamma}\sum_{j\neq i}\tilde{S}_j - \frac{(I-1)(1-\gamma)\eta}{\gamma})}{P_0 + IP_1}$$

will yield the most accurate posterior.

However, Credulous Bayesianism may operate in this setting as well, and in this case we assume that it causes people to underestimate the extent to which others have skewed their opinions to conform to the community norm. Instead, Credulous Bayesians assume that other people are using a reporting rule of $\tilde{S}_i = \lambda\gamma S_i + (1-\lambda\gamma)\eta$, where $\lambda \in [1, \frac{1}{\gamma}]$, which nests the two extremes of perfectly correcting for conformity and treating everyone else as being a completely honest reporter of their private signal. This changes the formula for the posterior belief to

$$\frac{P_1(S_i + \frac{1}{\lambda\gamma}\sum_{j\neq i}\tilde{S}_j - \frac{(I-1)(1-\lambda\gamma)\eta}{\lambda\gamma})}{P_0 + IP_1}.$$

In this case, there is no longer homogeneity within a group, because indi-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

viduals see their own signals accurately but incorrectly treat the signals of those around them. One interpretation of this assumption is that people suffer from a lack of higher order reasoning by failing to consider the motives of those around them. With this formula, Proposition 6 follows:

> *Proposition 6:*
>
> If $\eta > 0$ and $\lambda > 1$ then posterior beliefs will be biased upwards, and this bias increases with λ, η, I and P_1 and falls with P_0.
>
> If η is randomly distributed across groups, with mean 0 and variance $Var(\eta)$, then the variance of posterior beliefs is rising with λ if and only if $(I-1)(\lambda-1)Var(\eta) > \frac{I+\lambda-1}{P_0} + \frac{1}{P_1}$. The within-group (i.e. conditioning on D and η) variance in posterior beliefs is always declining with λ. The variance of the error in posterior beliefs is always rising with λ.

77 Proposition 6 shows how conformism in stating beliefs will affect Credulous Bayesians. A norm of stating a belief of η will cause biased posteriors if and only if $\lambda > 1$, so that people don't sufficiently correct for the fact that statements are conforming to a norm. The degree of that bias increases with the extent that people don't adequately control for conformism in people's statements (λ) and the magnitude of the norm (η).

78 Larger groups will have more bias because individuals base their beliefs more on the opinions of others and less on their own private signal. Credulous Bayesianism doesn't cause any error in an individual's private signal, but it does cause people incorrectly to assess the impact of their neighbor's views. As group sizes increase, there is a larger range of outside voices for people to misinterpret.

79 The amount of bias is also increasing with the variance of the true value (D) and declining with the variance of the noise terms. The connection between the bias and these two variance terms occurs because people weight the signals that they hear more highly when the variance in the underlying parameter is higher and the variance in the noise term is lower. As people weight the messages that they hear more heavily, the impact of Credulous Bayesianism naturally increases, because they are generally weighting these signals incorrectly.

80 Proposition 6 also tells us that the variance of the error in posterior beliefs is always rising with λ. When people get worse at correcting for con-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

formism, their beliefs become less accurate. Credulous Bayesianism causes an increase in the uniformity of beliefs within a group, as people conform in their statements to the group-wide norm, and those statements then further decrease the heterogeneity of beliefs within that group. (Recall the finding to this effect in the Colorado experiment.)

If the variance of norms across groups is sufficiently high, the combination of Credulous Bayesianism and conformism can actually increase the heterogeneity of posterior beliefs across groups and within the population as a whole. When the variance of norms is low, then Credulous Bayesians weight other people's opinions less, since people perceive less of a need to inflate the perceived difference between the actual statement and the norm. When the variance of norms is high, then Credulous Bayesianism makes these norms extremely powerful, since people believe that they receive information from their peers, while they are really just hearing statements that reflect the prevailing norm within their group. When the variance of norms is high, then Credulous Bayesianism acts mainly to cause people to inflate the importance of these norms. In this case, we should expect to see increased conformity within groups of Credulous Bayesians and increased heterogeneity across groups and across the population as a whole—what has been called "second-order diversity" (Gerken 2004).

The parameter γ does not appear in Proposition 6 because it is not conformism that matters but rather the extent to which people don't correct for that conformism, which is captured in the λ term. The degree of conformism does matter, however, if we assume that individuals completely fail to correct for the tendency to conform, and set $\lambda = \frac{1}{\gamma}$. In that case γ matters because it affects λ and Corollary 1 follows:

> *Corollary 1:*
>
> If people are naïve Bayesians who completely fail to correct for the tendency of statements to conform to the norm, then as γ falls and conformism rises, the positive bias of the posterior increases if $\eta > 0$, the variance of the error in posterior beliefs rises and the variance of beliefs within groups declines. The variance of posterior beliefs falls with γ if and only if $(l-1)(1-\gamma)$ $Var(\eta) > \frac{1+\gamma(l-1)}{P_0} + \frac{\gamma}{P_1}$.

If people are completely naïve Bayesians, then increased conformism in-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

creases the degree of error and bias in beliefs. As before, heterogeneity within groups will decline with the amount of conformism. Heterogeneity across people and groups can rise if the variation in the norm across society as a whole is sufficiently high.

84 These results can also help us understand experimental results showing that if members of the group think that they have a shared identity, and a high degree of solidarity, there will be heightened polarization. When people are placed in groups that emphasize some shared attribute, polarization is increased (Abrams *et al.* 1990, 112). A sense of "group belongingness" — of membership in a single group with salient shared characteristics — predictably affects the extent of polarization.[6]

85 A revealing experiment, fitting closely with the account we are offering here, attempted to investigate the effects of group identification on polarization (Spears, Lee & Lee 1990). Some subjects were given instructions in which group membership was made salient (the "group immersion" condition), whereas others were not (the "individual" condition). For example, subjects in the group immersion condition were told that their group consisted solely of first-year psychology students and that they were being tested as group members rather than as individuals. The relevant issues involved affirmative action, government subsidies for the theatre, privatization of nationalized industries, and phasing out nuclear power plants. Polarization generally occurred, but it was greater when group identity was emphasized. This experiment shows that polarization is highly likely to occur, and to be most extreme, when group membership is made salient.

86 In the context of the model, increases in λ and γ (parameters which capture the degree to which people make conformist statements and accept conformist statements as truth), can be interpreted as increases in the strength of group membership. Increasing the sense of solidarity presumably makes people more likely to trust one another, that is, to have a high value of λ, and possibly less likely to lie to each other, which would cause γ to rise. Higher values of both of these parameters will always increase the conformity within the group. As long as the variance of norms across groups is sufficiently high, increases in these parameters will also cause the

6 In the same vein, physical spacing tends to reduce polarization; a sense of common fate and intragroup similarity tend to increase it, as does the introduction of a rival "outgroup."

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

polarization of the groups to increase. This can explain the connection between group identity and group polarization.

In sum, conformism increases extremism and error when people are Credulous Bayesians. One way to distinguish between the conformist and rational models is to look at the impact of private and public communication. In the conformism model, public communications would look quite different from private communications, as we expect people to express the community norm. Moreover, individuals would be swayed by these conformist public statements. In a more rational model, individuals would put less weight on public statements than on private statements, when they know that people speak in ways that conform to a group norm.

7. PERSUASION

In this section, we explore the impact of self-interested persuaders on Credulous Bayesians. Our goal here is to show that Credulous Bayesianism can provide a model that makes sense of persuasive behavior, which is harder to rationalize with a perfectly rational model. The core idea of Credulous Bayesianism, which is that people pay too much attention to the opinions of others, also predicts that persuasion will matter and that we will see plenty of attempts at persuasion, even in circumstances in which fully rational people would not be much moved.

This model follows along the lines of Mullainathan, Schwarzstein and Shleifer (2007), who look at persuasion and coarse, categorical thinking. We now assume that statements are motivated not by a simple desire for conformism, but by a desire to influence an outcome, such as sentence length, environmental policy, guilt or innocence, or purchasing patterns. We assume that one or more decision-makers will choose an outcome that is equal to their posterior assessment of *D*. In this case, we assume that the decision-makers have no independent knowledge of the true value of *D*, other than its mean of zero and variance of $\frac{1}{P_0}$. All information beyond that comes from the statements of *I* other individuals who do have private signals with idiosyncratic error terms of variance $\frac{1}{P_1}$. To keep things simple, we return to the assumption of Section 3 that the private signals have independent error terms.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

90 As before, these other individuals have a taste for telling the truth, but because they are attempting to persuade, they also go in a particular direction. We capture these assumptions by assuming that these people maximize the expectation of $\pi\hat{D} - (\widetilde{S}_i - S_i)^2$ where π differs across the population and reflects the heterogeneous objectives of different actors and $\hat{D}$ denotes the chosen outcome. One interpretation of this model is that signals are being produced by two lawyers, one of which has a value of π of one and the other of which has a value of π of minus one. The two lawyers are both trying to persuade the judge and jury of the truth of their view of the case. They are constrained somewhat by the truth, but send signals that are biased towards their side of the case.

91 We constrain the decision-maker to use a linear updating rule so that posterior beliefs equal $\beta\Sigma_i\widetilde{S}_i - \Sigma_i K_i$, where β are endogeneously determined weights that will be discussed later and K_i is a constant that may be person specific but that is independent of the reported signal. Given this assumption, the individual's choice of reported signal will satisfy: $\widetilde{S}_i = S_i + \frac{0.5\beta\pi}{\gamma}$. In this case, the existence of incentives creates an additive error term which surrounds the signal.

92 We assume that the decision-maker does not know the values of $\frac{0.5\pi}{\gamma}$, but has an opinion about the distribution of this variable, which we denote μ_i. The variable is normally distributed with variance of $\frac{1}{P_\mu}$. If the true mean of this variable is zero, optimal signal extraction means that the decision-maker will set his estimate of the average value of K_i equal to zero if the individual cannot distinguish the motives of the speakers. If we assume that the decision-maker cannot commit to the weights that will be used ex post to form the posterior and the judgment, then the weights that minimize the variance of the error in the decision-maker's posterior belief will satisfy the equation

$$\frac{1}{P_0} = \frac{\beta I}{P_0} + \frac{\beta}{P_1} + \frac{2\beta^3}{P_\mu},$$

and we denote that value β^*.

93 In this case, Credulous Bayesianism will cause the decision-maker to underestimate the true heterogeneity of bias in the population by thinking that the variance of $\frac{0.5\pi}{\gamma}$ is $\frac{\lambda}{P_\mu}$. This assumption can be thought of as

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

a naïve trust in people's honesty or, again, as a failure to engage in higher-order reasoning that would lead to the view that people slant their statements to achieve an end. In this case, the optimal ex post value of $\beta^*(\lambda)$ satisfies

$$\frac{1}{P_0} = \frac{\beta I}{P_0} + \frac{\beta}{P_1} + \frac{2\lambda\beta^3}{P_\mu}.$$

Differentiating this condition implies that a higher value of λ causes the decision-maker to be more skeptical about the signals that are reported. We then prove in the appendix that:

Proposition 7:

(a) As λ rises, the variance of the decision-maker's posterior belief and the variance of the error associated with that beliefs when $\lambda < 1$. As λ increases, the decision-maker believes that the variance of the error in his posterior beliefs also increases.

(b) If the decision-maker believes that the expectation of $\frac{0.5\pi}{\gamma}$ is zero, but in the population this is not actually the case, then the expected error term in the posterior belief will be increasing with the mean of π and with the covariance of π and $\frac{1}{\gamma}$. The expected error will increase with the mean of $\frac{1}{\gamma}$ if and only if the mean of π is positive.

The proposition shows that increases in λ, the degree of cynicism about the motives of persuaders, create less heterogeneity in posterior beliefs and more accurate beliefs. This occurs for two reasons. First, greater skepticism means that the decision-maker tends to ignore the attempts at persuasion and sticks with his ex ante beliefs. Second, since the decision-maker is less susceptible to persuasion, the persuaders put less effort into misleading statements. Both effects together mean that the decision-maker's opinions are closer to his more accurate, and less variable, priors. Since persuasion creates both error and variability, less persuasion reduces both variability and error.

As in the case of other errors that come from the Credulous Bayesianism, incorrect beliefs have the effect of making posteriors more erroneous but increasing the confidence with which people hold to those erroneous beliefs. In this case, naïve decision-makers think that they are far more accurate than cynical decision-makers whose beliefs hew more closely to reality.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

96 Part (b) of the proposition discusses the ex post bias of decision-maker beliefs given that the mean of $\frac{0.5\pi}{\gamma}$ is not zero. In that case, increases in the mean value of π, the average incentive to persuade the decision-maker, will cause the bias to increase. Increases in the mean value of $\frac{1}{\gamma}$ which captures the average willingness to lie, will cause the extent of the bias to increase
97 if the mean of π is positive or decrease if the mean of π is negative. The willingness to lie tends to exacerbate the biases that come from an uneven distribution of incentives.

Finally, the covariance between the incentives to lie and the willingness to lie is also important. When these two things are more likely to go together, then the bias in posterior beliefs will increase. These results suggest that it isn't the presence of incentives to lie that cause biased decisions. The problem comes when those incentives are unevenly distributed or correlated with the ability to lie.

98 So far we have assumed that the decision-maker has no idea about the biases that may afflict particular informants. Now we take the opposite assumption by assuming that the decision-maker has an assessment of μ_i for each individual equal to $\lambda_i\mu_i$. We assume that the decision-maker believes that there is no error in his assessment of the bias. In this case, the decision-maker sets

99

$$\beta = \frac{P_1}{P_0 + IP_1} \text{ and } K_i = \left(\frac{P_1}{P_0 + IP_1}\right)^2 \lambda_i \mu_i$$

for each one of the speakers. With these formulae, it follows that:

> *Proposition 8:*
> If $\lambda_i = \lambda$ for all i, then the variance of posterior beliefs and the variance of the error in posterior beliefs declines with λ if λ is less than one. The decision-maker's perceived variance of the error in his posterior declines with λ for λ between $\frac{1}{3}$ and 1 and increases with λ for all other values. The expected level of bias is increasing in $Cov(\pi_i, \frac{1}{\gamma_i})$, $Cov(1-\lambda_i, \mu_i)$, $E(\pi_i)$ and I. It is increasing with $E(\frac{1}{\gamma_i})$ and decreasing with the average of λ_i if and only if $E(\pi_i)$ is positive.

The proposition first makes the point that if the same incorrect level of adjustment for incentives is applied to everyone, we should expect the vari-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

ance of the posterior and the error around the posterior to fall as λ rises. Both effects occur because higher values of λ purge the statements of their biases and those biases create both excess variance and error.

The expected level of bias in the posterior belief, and the decision, is a function of the mean level of $(1-\lambda_i)\mu_i$ — the uncorrected bias terms in the individual statements. This mean level will be increasing in the covariance between the extent that the decision-maker fails to correct fully for bias and the degree of bias. If λ_i is constant across i, then this covariance term equals zero. If the decision-maker is more likely to be appropriately cynical towards only one side of the debate, then there will be biased decision-making.

Increases in the mean value of π_i will increase the bias, because this implies that incentives are stacked more strongly on one side of the debate than on the other. If the mean value of π_i is positive, then increasing the value of $E(\frac{1}{\gamma_i})$, which effectively means decreasing the cost of lying, will cause the level of bias to increase. As before, covariance, either positive or negative, between incentives and the ability to lie will increase the expected level of bias.

Increases in I will cause the bias to go up because as the number of people increases, the decision-maker puts more weight on the biased views of those people. This last result assumes that the overall mean level of bias is independent of I, which might not be true in practice. This result points again to the possibility that crowds might be more foolish than individuals.

While these results may seem unsurprising in light of the literature on persuasion, the model provides a simple framework that illustrates the similarities between persuasion and other forms of group communication. If individuals are insufficiently skeptical towards their friends, then we should not be surprised that they are also insufficiently skeptical towards advertisers. Both phenomena reflect the same tendency to put excessive faith in the stated opinions of others.

We can also use this framework to make sense of an experiment designed to see how group polarization might be dampened (Abrams *et al.* 1990, 112). The experiment involved the creation of four-person groups. On the basis of pretesting, these groups were known to include equal numbers of persons on two sides of political issues — whether smoking should be banned in public places, whether sex discrimination is a thing of the past, whether censorship

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

of material for adults infringes on human liberties, and so on. Judgments were registered on a scale running from +4 (strong agreement) to 0 (neutral) to –4 (strong disagreement). In half of the cases (the "uncategorized condition"), subjects were not made aware that the group consisted of equally divided subgroups in pretests. In the other half (the "categorized condition"), subjects were told that they would find a sharp division in their group, which consisted of equally divided subgroups. They were also informed who was in which group and told that they should sit around the table so that one subgroup was on one side facing the other group.

In the uncategorized condition, discussion generally led to a dramatic reduction in the gap between the two sides, thus producing a convergence of opinion toward the middle of the two opposing positions (a mean of 3.40 scale points, on the scale of +4 to –4). But things were very different in the categorized condition. Here the shift toward the median was much less pronounced, and frequently there was barely any shift at all (a mean of 1.68
105 scale points). In short, calling attention to group membership made people far less likely to shift in directions urged by people from different groups.

In our model, the distinction between categorized and uncategorized groups can be seen as comparing a scenario in which individuals know the value of $\frac{0.5\pi}{\gamma}$ and a scenario in which they do not. We will pare the experiment down to its essentials by assuming that there are exactly two people in each group, each of whom acts as both a decision-maker and a persuader. We also must assume that individuals have their own private signals. In both sce-
106 narios, one member of the dyad has a value of $\frac{0.5\pi}{\gamma}$ equal to 1 and the other has a value of –1. We also assume that the signal of the individual for whom $\frac{0.5\pi}{\gamma}$ equals 1 is higher than the individual for whom $\frac{0.5\pi}{\gamma}$ equals –1.

In the first scenario, both individuals believe that their partner is randomly drawn from the population both in their signal and in their value of $\frac{0.5\pi}{\gamma}$. They also still believe that the population mean of $\frac{0.5\pi}{\gamma}$ is zero. In the second scenario, the partners know the value of $\frac{0.5\pi}{\gamma}$ for their partner. In both scenarios, people report signals equal to $\widetilde{S}_i = S_i + \frac{0.5\beta\pi}{\gamma}$, where the value of β reflects the signal extraction formula used by the listened. Of
107 course, the value of β differs between the two scenarios.

In the first scenario, when $\frac{0.5\pi}{\gamma}$ is not known, and people are Credulous Bayesians, the value of β solves

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

$$\frac{1}{P_0+P_1}=\frac{(P_0+2P_1)\beta}{(P_0+P_1)P_1}+\frac{2\lambda^2\beta^3}{P_\mu},$$

and we denote that value β^*. The weight that the individuals put on their own signal equals $\frac{P_1(1-\beta^*)}{P_0+P_1}$. The posterior belief for individual i (for whom $\frac{0.5\pi}{\gamma}$ equals one) will be

$$\frac{P_1(1-\beta^*)}{P_0+P_1}S_i+\beta^*S_j-\beta^{*2}.$$

The difference between the two individuals opinions will equal

$$\frac{(S_i-S_j)(P_1(1-2\beta^*)-P_0\beta^*)}{P_0+P_1}-2\beta^{*2}.$$

As β^* falls with λ, the gap between the two participants will rise as people become more cynical, so that more trust is associated with more homogeneity in the non-characterization treatment.

In the case where people know each other's values of $\frac{0.5\pi}{\gamma}$, then perfect Bayesians would correct completely for these incentives. Ex post everyone would know the right answer and they would agree on their opinions. Clearly, perfect Bayesian learning cannot explain the observed non-agreement in the categorization version of the experiment.

However, a particular form of Credulous Bayesianism can better match the facts. If individuals completely discounted information from people who are known to have opposing views and who are shading their statements accordingly, then this would ensure a complete failure to reach any sort of consensus. In this setting, Credulous Bayesianism would mean that we discretely categorize people into friends and enemies and completely ignore the statements of enemies.

This type of extreme categorization might enable us to make sense of the sharp differences in beliefs that we observe across groups. The preceding models stressed that these differences could reflect Credulous Bayesianism, which results in a tendency to take neighbors' statements too seriously. However, in those models individuals were unaware of views held outside of the group. In the real world, people are often well aware that there are others who hold differing opinions. In either a standard Bayesian model or a model where people are Credulous Bayesians, those differing opinions should cause a substantial divergence across groups. In such a Bayes-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

ian model, American Christians should reduce their faith in God because Indians have a different belief system. Palestinians who believe that Mossad destroyed the World Trade Center should moderate their views because they know that Americans do not share that view. If people thought that the views expressed by other groups were completely untrustworthy and shaped more by incentives to lie than by information, then we might be able to understand this failure of beliefs to converge.

8. COGNITIVE DIVERSITY AND INSTITUTIONAL DESIGN

If people are acting as Bayesians, they will end up both more unified and more extreme as a result of group discussions. The result may be nothing to deplore. If group members begin with the thought that people likely have one heart and two kidneys, or that it is probably negligent to drive over 80 miles per hour in a crowded area near a school, there is no problem if discussion leads them to become more firmly committed to these beliefs and
112 more unified in holding them. The same idea expresses the ideal conception of jury deliberation, as the exchange of information is supposed to ensure unanimity on a proposition that is true; and indeed juries typically polarize in criminal cases (Brown 1986). The notion of deliberative democracy has similar foundations. If the initial distribution of information is adequate, nothing is wrong with a situation in which participants in democratic deliberation become more extreme in their commitment to a certain outcome or course of action.

If, on the other hand, people are Credulous Bayesians, and overreacting to the actual or perceived views of others, they may end up making major mistakes. As we have seen, large groups may do worse, not better, than small ones. With respect to politics, people may accept some view that is clearly inconsistent with the facts; widespread commitments to implausible conspiracy theories, or to preposterous accounts of what "really" underlies
113 some natural or social phenomenon, can be understood in this light. Actual behavior may be adversely affected as well — as, for example, when people falsely believe that they are at risk and take unjustified precautions, or falsely believe that they are safe and fail to take protective measures.

Of course it is not possible to move directly from an understanding of how social learning occurs to any particular set of institutional reforms.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Much needs to be known about the particular context, including the particular form of Credulous Bayesianism. If people fail sufficiently to discount the self-interested incentives of speakers, as in the case of political advertising, the solution may be different from what it should be if people fail sufficiently to discount for the skewed nature of the distribution of information within their group. In some groups, people may be good Bayesians and engage in appropriate discounting. Perhaps expert institutions are able to do exactly that. In other groups, diversity is desirable, but it also has significant costs in terms of (for example) increased acrimony, social loafing, and greater difficulty in reaching any decision at all. The benefits of error reduction may be lower than those costs.

With these disclaimers, we explore some possible implications here for independent regulatory agencies and federal appellate courts — two institutions that are extremely important and that are objects of considerable public debate. Our goal is not to make any particular normative recommendation, but to obtain a better understanding of current practices requiring diversity (in the case of administrative agencies) and current debates over the issue of diversity (in the case of federal appellate courts). We also offer a brief note on media policy. In particular, we are interested in the relationship between our claims and debates, past and present, over the "fairness doctrine" long imposed and now abandoned by the Federal Communications Commission.

8.1. Independent Regulatory Commissions

A great deal of national policy is established by the so-called independent regulatory commissions, such as the Federal Communications Commission, the Federal Trade Commission, the Securities and Exchange Commission, and the National Labor Relations Board. These agencies typically consist of five members, who are appointed by the president (with the advice and consent of the Senate), serve for specified terms (usually seven years), and make decisions by majority vote. Because of the immense importance of their decisions, any Democratic president would much like to be able to ensure that the commissions consist entirely or almost entirely of Democratic appointees; Republican presidents would certainly like to shift policy in their preferred directions by ensuring domination by Republican appointees.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Under existing law, however, presidential control of the commissions is sharply constrained, for no more than a bare majority can be from a single political party. Congress has explicitly so required (typically saying, "not more than three of the Commissioners shall be members of the same political party"),[7] and indeed this has become the standard pattern for the independent agencies. Hence, for example, the National Labor Relations
117 Board and the Federal Communications Commission must consist either of three Republicans and two Democrats or of three Democrats and two Republicans. From the standpoint of the president, a particular problem arises in a time of transition from one administration to the next. A Democratic president, for example, is often disturbed to learn that agencies entrusted with implementing legislation policy will be composed of at least two Republicans (appointed by his predecessor).

It should be clear that the requirement of bipartisan composition operates as a constraint on group polarization and extreme movements. Five Democratic appointees to the NLRB, for example, might well lead labor law in dramatic new directions. As we have seen, such movements could operate through rational Bayesianism. Perhaps a president would like to choose five people with extensive experience in labor-management relations, specializing in marshalling evidence and arguments in support of labor unions. Perhaps the five Democratic appointees could learn from one another in a way that produces a consensus on some position that, while
118 extreme in light of existing law, is sensible as a matter of policy. Or perhaps the movements could occur as a result of Credulous Bayesianism. NLRB commissioners might well discount the extent to which the information that they hold is shared by all, or the extent to which important views are missing, or the extent to which some of them are signaling a position that conforms to the perceived group norm. Such signaling can occur within expert groups as well as within groups of nonspecialists. If people are Credulous Bayesians, then the presence of two Republican appointees constrains the relevant movements and ensures that significant counterarguments will be offered. To this extent, bipartisan membership might serve to limit unwarrantedly extreme changes in regulatory policy.

7 *See* 47 U.S.C. §154(b)(5) (Federal Communications Commission); 15 U.S.C. §41 (Federal Trade Commission).

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

We can now understand how requirements of bipartisan membership might reasonably be debated. In the abstract, it is not clear that any particular Congress would want to prevent relative extremism. A Democratic-majority Congress, and the groups who support its members, might well believe that an all-Democratic NLRB would have a better understanding of national labor policy; perhaps the rulings of such an NLRB would be a more faithful agent of that particular legislature. If Democratic members 119 are perfect Bayesians, and if an all-Democratic NLRB contained the optimal range of information, so that Republican appointees would add confusion and falsehood, bipartisan composition would be hard to justify. (No legislator believes that the NLRB should have communist or anarchist members.) But if many members of Congress believe that stability is desirable over time, and if most of them want to check unjustified movements produced by Credulous Bayesianism, legislators, and the diverse groups who pressure them, might be able to reach a consensus on bipartisan membership as the best means to their ends. Bipartisan membership might turn out to represent a stable kind of arms control agreement, in which members of both parties are willing to relinquish the possibility of extreme movements in their preferred direction in return for assurance against extreme movements the other way.[8] And some legislators, and outside observers, might be willing to defend the current situation as reflecting an intuitive awareness of the consequences of Credulous Bayesianism.

8.2. Federal Appellate Courts

Do similar considerations apply to the federal judiciary? At first glance, the judiciary is quite different, because many people believe that it is not supposed to make policy at all. And indeed, judges are supposed to be specialists in assessing both (legally relevant) facts and law, and hence the idea of bipartisan membership might seem jarring. But the evidence suggests a more complicated picture. Note first that judicial panels consist of three judges, and assignment to three-judge panels is random. This means that there are many DDD panels, many RRR panels, many RDD panels, and 120 many RRD panels. As our analysis would predict, extreme movements are

8 For informative discussion, involving alternating incumbency as a way of constraining extremism on the bench, *see* J. Mark Ramseyer & Eric B. Rasmussen (2003).

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

shown by DDD and RRR panels, in the sense that judges, on such panels, are especially likely to vote in line with ideological stereotypes.

We have referred to this point in general terms; now consider a few examples (Sunstein *et al.* 2006). On all-Republican panels, Republican appointees vote for gay rights 14 percent of the time; on all-Democratic panels, Democratic appointees vote for gay rights 100 percent of the time. On
121 all-Republican panels, Republican appointees vote to validate affirmative action programs 34 percent of the time; on all-Democratic panels, Democratic appointees vote to validate such programs 83 percent of the time. On all-Republican panels, Republican appointees vote in favor of women in sex discrimination cases 30 percent of the time; on all-Democratic panels, Democratic appointees vote in favor of women in sex discrimination cases 76 percent of the time. On all-Republican panels, Republican appointees vote for disabled people in cases brought under the Americans with Disabilities Act 17 percent of the time; on all-Democratic panels, Democratic appointees vote for disabled people in such cases 50 percent of the time. In cases brought under the National Environmental Policy Act, Republican appointees on all-Republican panels vote for environmental plaintiffs 20 percent of the time; in such cases, Democratic appointees on all-Democratic panels vote for environmental plaintiffs 71 percent of the time.

By contrast, both Republican and Democratic appointees show far more moderation when they sit on panels containing at least one appointee nominated by a president of the opposing political party. On mixed panels, Republican appointees are much more liberal than they are on unified Republican panels; Democratic appointees show precisely the same effects. In affirmative action cases, Republican appointees show a 69 percent liberal voting rate on RDD panels (far above the 34 percent rate on RRR panels); in such cases,
122 Democratic appointees show a 60 percent liberal voting rate on DRR panels (far below the 83 percent rate on DDD panels). In sex discrimination cases, Republican appointees show a 44 percent liberal voting rate on RDD panels (far above the 30 percent rate on RRR panels); in such cases, Democratic appointees show 58 percent liberal voting rate on DRR panels (far below the 76 percent rate on DDD panels). In some domains, the difference between Democratic appointees and Republican appointees is small or even nonexistent on mixed panels; it emerges only when we compare Rs on RRR panels

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

to Ds on DDD panels. In nearly all areas, the difference between Republican and Democratic appointees is sharpest on DDD and RRR panels; sitting with like-minded judges appear to create significant polarization. We have said that the NLRB must have bipartisan membership, but of course appellate panels that review the NLRB need not; and the results of appellate review of NLRB decisions are very different depending on whether the panel is RRR or DDD (*see* Miles & Sunstein 2008).

These patterns are consistent with perfect Bayesianism. On a DDD panel, Democratic appointees will hear different conclusions and different arguments from what they hear on a DRR panel. And if DDD panels contain all the arguments that it is useful to hear, nothing is amiss. But it is at least possible that in some cases, such appointees are not receiving new information at all or that they should discount the relevant arguments by taking account of the sources. In short, Credulous Bayesianism may well be at work, even on the federal bench. Some of the time, we speculate, judges may well be acting as if agreement from other judges supplies additional information, when it does not. A great deal of additional work would be needed to understanding the precise mechanisms here; we could imagine experiments in which judicial deliberations were recorded and analyzed. But in support of our speculation, consider Judge Richard A. Posner's report that serious deliberation, and the careful exchange of information and reasons, are rare among three-judge panels (*see* Posner, 2008). If this is so, it is reasonable to believe that Credulous Bayesianism helps to account for the observed patterns. And at least some of the time, it is also reasonable to believe that in ideologically contested cases, the greater moderation of DDR or RRD panels is influenced by the existence of competing conclusions and even arguments.

The purpose of full or "en banc" review, on the courts of appeals, is to correct errors on the part of three-judge panels. If that is the purpose, and if we do not believe that Ds or Rs have a monopoly on information or wisdom, a relatively uncontroversial implication is that a warning flag should be raised whenever a unified panel goes far in the ideologically predictable direction. That warning flag might justify closer consideration of en banc review. In the most important cases, the warning flag might also be relevant to the Supreme Court's decision whether to take review, which is also designed to correct

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

errors on the part of lower courts. It would seem quite sensible for the Supreme Court to consider, as a relevant factor, whether the decision it is being asked to review was decided by a unified or mixed panel. If a DDD panel has ruled in favor of an affirmative action program, or if an RRR panel has ruled against environmentalists challenging a federal regulation, there is particular reason to attend to an argument that the panel has erred. In fact we are willing to hypothesize that the Court's reviewing practice is implicitly responsive to this consideration, and that the Court is distinctly likely to grant review in ideologically contested cases resolved by a DDD or RRR panel. It would be valuable to test this hypothesis empirically.

A possible counterargument would be that while the political party of the appointing president is a proxy for ideology, the proxy is crude: Some Republican appointees are more liberal, in general or in particular areas, than some Democratic appointees. A more fine-grained approach, attentive to the value of diversity, would inquire directly into the established voting tendencies of various judges, not into the political party of the appointing president. But it is not simple to operationalize the more fine-
125 grained measures, even though they exist; and during judges' early years on the bench, judicial records are too spare to permit easy characterizations. The political party of the appointing president may be the best way to combine (adequate) accuracy with ease of administration.

A much more controversial implication is that in the most difficult and ideologically charged cases, those who seek to avoid the effects of group polarization should consider efforts to create diverse judicial panels, as in the context of the NLRB and the FCC. (Of course any proposal in favor of balanced panels would be feasible only in circuits that have significant numbers of both Ds and Rs.) This implication is controversial because the judiciary is not understood as a policymaking institution, because such an approach
126 might cement judicial self-identification in political terms, and because efforts to ensure ideological diversity might well be taken as inconsistent with the commitment to judicial neutrality. But the discussion here suggests that judges are policymakers of a distinctive kind, and that in principle, the argument for diversity, as a means of counteracting Credulous Bayesianism and hence group polarization, is not significantly different from the argument in the context of the independent regulatory commissions. Recall here that while the NLRB must be DDDRR or RRRDD, reviewing courts are not simi-

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

larly constrained, and that the ultimate fate of NLRB decisions and hence national labor law, even in the most important domains, will often be radically different if the reviewing court is RRR or DDD. By contrast, appellate panels are far more moderate if they are RRD or RDD.

Of course it is also true that judicial diversity is constrained by a number of other factors, most obviously the fact that judicial panels are composed entirely of lawyers. We put to one side the interesting question whether this lack of diversity causes significant problems in terms of our central argument here.

8.3. Media Policy and Diversity

For many years, the Federal Communications Commission imposed the 127
"fairness doctrine," which, in brief summary, required radio and television broadcasters to cover issues of public controversy and to allow presentations by competing sides (Sunstein 1993). Under the fairness doctrine, it would be unacceptable for a television station to offer the "liberal" position on all issues, without permitting alternative positions to have their say. The fairness doctrine was of course highly controversial, in part because of evident difficulties in administration, and indeed it was challenged in 128
the Supreme Court on the ground that it abridged the free speech rights of broadcasters (*Red Lion Broadcasting Co. v. FCC*[9]). In the view of the challengers, the government should not be allowed to force them to present certain positions on the stations that they owned.

For our purposes, the Court's response was of special interest. The Court emphasized that the "rights of listeners and viewers," rather than the rights of broadcasters, should be taken as paramount. In the Court's views, listeners and viewers had something like a "right" to be exposed to competing positions, and a single set of presentations, from a single point of view, would violate that "right." This claim might seem puzzling in the abstract; whether or not is it correct, it is far easier to understand the Court's concern in light of the arguments we have offered. Credulous Bayesians might fail 129
to discount the partiality or bias of a station that offers a particular point of view; they might treat the presentation that is offered as relevantly representative even though it is not. Without the fairness doctrine, there is a risk

9 395 U.S. 367 (1969).

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

that people will live in echo chambers, or information cocoons, in which they end up more confident and more extreme simply because they are listening to the same point of view.

We do not mean here to explore the greatly contested question whether the fairness doctrine was a good idea, even in its time of great scarcity of broadcasters; the point is only that the doctrine, and the Court's decision to uphold it, may well be understood as reflecting an intuitive understanding of Credulous Bayesianism. Ordinary listeners and viewers, hearing a station
130 that repeatedly offers a single perspective, might not sufficiently discount the points of view that they are hearing.

In the modern era, the FCC has mostly repealed the doctrine's requirements, largely on the theory that with so many options and outlets, people are able to have access to an exceedingly wide range of information and opinions. Nothing said here demonstrates that this conclusion was wrong. But the remaining problem, signaled by our analysis, is that if people are engaged in a degree of self-sorting, so that they select points of view with which they antecedently agree, they might well be moved in the direction of
131 extremism precisely because of the operation of Credulous Bayesianism.

To the extent that there is a high degree of self-sorting, the communications market may reveal a fully voluntary version of the Colorado experiment. The market is likely to have some similar dynamics, underpinned by Credulous Bayesianism, producing both polarization and confidence. In fact we would predict that a fully open communications market, with ideologically identifiable sources and with voluntary sorting, would, for many people, replicate the results of that experiment.

132

9. CONCLUSION: THE WISDOM AND THE FOLLY OF CROWDS

We have argued here that extreme movements can be a product of rational updating, as people respond to the information and the arguments offered by others. To this extent, polarized opinions need not reflect any kind of bias or irrationality on the part of those whose opinions have been rendered more extreme. To the extent that people are responding rationally to new information, more confident and more extreme groups may well be wise. The "wisdom of crowds" is plausibly explained in this way.

133

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

At the same time, polarization may often be produced by Credulous Bayesianism, in which people treat the views of others as significantly more informative than they actually are. We have explored four possibilities. (1) Sometimes people's opinions have common sources, and hence the views of others add little. (2) Sometimes group members are not a random sample 134 of the population as a whole and the pre-deliberation distribution of views is biased. (3) Sometimes group members frame their views so as to curry favor or to avoid social sanctions. (4) Sometimes people have incentives to mislead. We have suggested that Credulous Bayesians give insufficient weight to these possibilities, in a way that can produce significant blunders. Errors by deliberating groups are frequently a product of these four phenomena. As a result of these errors, members of deliberating groups may well be less wise as well as more extreme than individuals. The folly of crowds is often a result.

An understanding of Credulous Bayesianism does not lead to any simple prescription for institutional design, but it does have important implications, and it helps explain a number of current practices and debates. Congress' decision to require bipartisan composition on the independent regulatory commissions, and the absence of a significant controversy over that decision, might be explained as responsive to an understanding of the risks associated with the possible movements that we have explored here. Much more controversially, we have suggested that an understanding of group polarization and Credulous Bayesianism helps to explain current calls for diversity on federal appellate panels. The question of appropriate media policy raises many complexities that we have not explored, but it is plain that many past and current debates are rooted in an intuitive awareness of the phenomenon of group polarization and an understanding that Credulous Bayesianism might lead both individuals and groups in unfortunate directions.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

REFERENCES

Abrams, Dominic, Margaret Wetherell, Sandra Cochrane, Michael A. Hogg & John C. Turner. 1990. Knowing What To Think By Knowing Who You Are: A Social Identity Approach to Norm Formation, Conformity, and Group Polarization, 29 *Brit. J. Soc. Psychol.* 97–119.

Asch, Solomon. 1955. Opinions and Social Pressure, 193 *Scientific American* 31–35.

Baron, Robert S., Sieg I. Hoppe, Bethany Brunsman, Barbara Linneweh & Diane Rogers. 1996. Social Corroboration and Opinion Extremity, 32(6) *J. Experimental Soc. Psychol.* 537–560.

Bernheim, B. Douglas. 1994. A Theory of Conformity, 102(5) *J. Pol. Econ.* 841–877.

Brown, Roger. 1986. *Social Psychology: The Second Edition*. New York: The Free Press.

Camerer, Colin F., Teck-Hua Ho & Kuan Chong. 2004. A cognitive hierarchy model of behavior in games, 119(3) *Q.J. Econ*. 861–898.

Cromwell, Paul, Alan Marks, James N. Olson & D'Aunn W. Avary. 1991. Group Effects on Decision–Making by Burglars, 69(2) *Psychol. Rep.* 579–588.

Crutchfield, R.S. 1955. Conformity and Character, 10 *Am. Psychol. J.* 191–198.

DeMarzo, Peter, Dimitri Vayanos & Jeffrey Zwiebel. 2003. Persuasion Bias, Social Influence and Uni–Dimensional Opinions, 118(3) *Q.J. Econ*. 909–968.

Eyster, Erik, & Matthew Rabin. 2005. Cursed Equilibrium, 73(5) *Econometrica* 1623–1672.

Getzkow, Matthew, & Jesse Shapiro. 2004. Media, Education, and Anti-Americanism in the Muslim World, 18(3) *J. Econ. Persp.* 117–133.

Gigerenzer, Gerd. 2007. *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking.

Glaeser, Edward L. 2004. Psychology and the Market, 94(2) *Am. Econ. Rev. Papers & Proc*. 408–413.

Glaeser, Edward L., & Bryce Ward. 2006. Myths and Realities of American Political Geography, 20(2) *J. Econ. Persp*. 119–144.

Habermas, Jurgen. 1998. *Between Facts and Norms*. Cambridge, Mass.: MIT.

Hong, Lawrence. 1978. Risky Shift and Cautious Shift: Some Direct Evidence on the Culture Value Theory, 41(4) *Soc. Psychol*. 342–346.

Kaplan, Martin F. 1977. Discussion Polarization Effects in a Modified Jury Decision Paradigm: Informational Influences, 40(4) *Sociometry* 262–271.

Kerr, Norbert, Robert MacCoun & Geoffrey P. Kramer. 1996. Bias in Judgment: Comparing Individuals and Groups. 103 *Psychol. Rev.* 687–705.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Krech, David, Richard S. Crutchfield & Egerton S. Ballachey. 1962. *The Individual In Society*. New York: McGraw Hill.

Lord, Charles G., Lee Ross & Mark R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence, 37 *J. Personality & Soc. Psychol.* 2098–2109.

Miles, Thomas, & Cass R. Sunstein. 2008. The Real World of Arbitrariness Review, 75 *U. Chi. L. Rev.* 761–813.

Morris, Stephen. 2001. Political Correctness, 109(2) *J. Pol. Econ.* 231–265.

Moscovici, Serge, & Marisa Zavalloni. 1969. The Group as a Polarizer of Attitudes, 12 *J. Personality & Soc. Psychol.* 125–135.

Mullainathan, Sendhil, Joshua Schwarzstein & Andre Shleifer. 2008. Coarse Thinking and Persuasion, *Q.J. Econ*.

Myers, David G. 1975. Discussion-Induced Attitude Polarization, 28(8) *Hum. Rel.* 699–714.

Myers, David G., & George D. Bishop. 1970. Discussion Effects on Racial Attitudes, 169 *Sci.* 778–789.

Myers, David G., & Martin F. Kaplan. 1976. Group-Induced Polarization in Simulated Juries, 2 *Personality & Soc. Psychol. Bull.* 63–66.

Page, Scott. 2006. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies.* Princeton: Princeton University Press.

Posner, Richard A. 2008. *How Judges Think*. Cambridge, Mass.: Harvard University Press.

Ramseyer, J. Mark, & Eric B. Rasmussen. 2003. *Measuring Judicial Independence*. Chicago: University of Chicago Press.

Ross, Lee, & Richard E. Nisbett. 1991. *The Person and the Situation*. New York: McGraw Hill.

Sargent, Thomas. 1979. *Macroeconomic Theory.* New York: Academic Press.

Schkade, David, Cass R. Sunstein & Reid Hastie. 2007. What Happened on Deliberation Day?, 95 *Cal. L. Rev.* 915–940.

Schkade, David, Cass R. Sunstein & Daniel Kahneman. 2000. Deliberating About Dollars: The Severity Shift, 100 *Colum. L. Rev.* 1139–1176.

Shapiro, Jesse. 2006. A Memory-Jamming Theory of Advertising. Mimeograph, University of Chicago.

Spears, Russell, Martin Lee & Stephen Lee. 1990. De-Individuation and Group Polarization in Computer-Mediated Communication, 29 *Brit. J Soc. Psychol.* 121–134.

Stoner, James A.F. 1961. A Comparison of Individual and Group Decision Involving Risk. Unpublished Master's Thesis, Massachusetts Institute of Technology. Available at: http://hdl.handle.net/1721.1/11330.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Sunstein, Cass R. 1993. *Democracy and the Problem of Free Speech*. New York: The Free Press.

———. 2003. *Why Societies Need Dissent*. Cambridge, Mass.: Harvard University Press.

Sunstein, Cass R., David Schkade & Lisa M. Ellman. 2004. Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation, 90(1) *Va. L. Rev.* 301–354.

Sunstein, Cass R., David Schkade, Lisa M. Ellman & Andres Sawicki. 2006. *Are Judges Political? An Empirical Investigation of the Federal Judiciary*. Washington: The Brookings Institution.

Surowiecki, James. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Little, Brown & Co.

Turner, John C., Michael A. Hogg, Penelope J. Oakes, Steven D. Reicher & Margaret S. Wetherell. 1987. *Rediscovering the Social Group:* A Self-Categorization Theory. New York: Blackwell.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

APPENDIX: PROOFS OF PROPOSITIONS

Proof of Proposition 1:

The derivative of the variance of ex post beliefs, or

$$\frac{IP_1(P_0(1+(I-1)\nu)+IP_1)}{P_0(P_0(1+(I-1)\lambda\nu)+IP_1)^2}$$

with respect to λ equals

$$-\frac{2I(I-1)\nu P_1(P_0(1+(I-1)\nu)+IP_1}{(P_0(1+(I-1)\lambda\nu)+IP_1)^3}<0.$$

The variance in the error of ex post beliefs is

$$\frac{(1-\nu)(I-1)\nu P_1 I}{(P_0(1+(I-1)\lambda\nu)+IP_1)^2}+\frac{1+(I-1)\lambda\nu}{P_0(1+(I-1)\lambda\nu)+IP_1}$$

and the derivative of this with respect to λ equals

$$-\frac{2P_0P_1I(I-1)^2(1-\lambda)\nu^2}{(P_0(1+(I-1)\lambda\nu)+IP_1)^3}<0.$$

The difference between the actual variance of the error term and the perceived variance of the error term is

$$\frac{(1-\lambda)(I-1)\nu P_1 I}{(P_0(1+(I-1)\lambda\nu)+IP_1)^2}$$

and the derivative of this with respect to λ equals:

$$-\frac{(I-1)\nu P_1 I\,(2(I-1)(1-\lambda)P_0\nu+P_0(1+(I-1)\lambda\nu)+IP_1)}{(P_0(1+(I-1)\lambda\nu)+IP_1)^3}<0.$$

The derivative of

$$\frac{(1-\lambda)(I-1)P_1 I}{(P_0(1+(I-1)\lambda\nu)+IP_1)^2}$$

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

with respect to ν equals

$$\frac{(P_0(1-(I-1)\lambda\nu)+IP_1)(1-\lambda)(I-1)\nu P_1 I}{(P_0(1+(I-1)\lambda\nu)+IP_1)^3}$$

which is positive if and only if

$$P_0+IP_1 > (I-1)\lambda\nu P_0.$$

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proof of Proposition 2:

The derivative of

$$\frac{(1-\nu)(I-1)\nu P_1 I}{(P_0(1+(I-1)\lambda\nu)+IP_1)^2}+\frac{1+(I-1)\lambda\nu}{P_0(1+(I-1)\lambda\nu)+IP_1}$$

$$\text{or}\quad \frac{(1+(I-1)\lambda\nu)^2 P_0+(1+(I-1)\nu)IP_1}{(P_0(1+(I-1)\lambda\nu)+IP_1)^2}$$

with respect to I equals

$$-\frac{P_0P_1[1+(I-1)\lambda\nu(\nu(1-2\lambda)+3)+(1-2I)\nu]+(1-\nu)\,IP_1{}^2}{(P_0(1+(I-1)\lambda\nu)+IP_1)^3}.$$

When $\lambda=1$, this equals

$$\frac{-(1-\lambda)P_1}{(P_0(1+(I-1)\nu)+IP_1)^2}$$

which is strictly negative, and since the derivative is continuous, it will remain negative when λ is high. When $\lambda=0$, the derivative equals

$$\frac{P_0P_1(2I\nu-1-\nu)-(1-\nu)IP_1{}^2}{(P_0+IP_1)^3}$$

which is strictly positive if and only if

$$\nu>\frac{P_0+IP_1}{P_0(2I-1)+IP_1{}^2}.$$

Again by continuity, the derivative will be positive when λ is sufficiently close to zero if this condition holds.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proof of Proposition 3:

The variance of ex post beliefs equals the variance of

$$\frac{P_1(\sum_i \eta_i + ID)}{P_0(1+(I-1)\lambda\psi)+IP_1}$$

which equals

$$\frac{P_1((I+I(I-1)\psi)P_0+I^2P_1)}{P_0(P_0(1+(I-1)\lambda\psi)+IP_1)^2}.$$

This variance is clearly declining with λ. The error in the beliefs is

$$\frac{P_1(\sum_i \eta_i)-P_0(1+(I-1)\lambda\psi)D}{P_0(1+(I-1)\lambda\psi)+IP_1}$$

and the variance of that is

$$\frac{P_1(I+I(I-1)\psi)+P_0(1+(I-1)\lambda\psi)^2}{(P_0(1+(I-1)\lambda\psi)+IP_1)^2}.$$

The derivative of that with respect to λ is

$$\frac{-(1-\lambda)2P_0P_1(I-1)^2I\psi^2}{(P_0(1+(I-1)\lambda\psi)+IP_1)^3}<0.$$

The derivative of the variance of ex post beliefs with respect to I is

$$\frac{P_0P_1[(1-\lambda\psi)+I\psi(1-\lambda)+\psi(I-1)(1-\lambda\psi)+I\frac{P_1}{P_0}(1+\psi-2\lambda\psi)]}{(P_0(1+(I-1)\lambda\psi)+IP_1)^3}>0.$$

The derivative of the variance of the error term with respect to I is

$$\frac{P_0P_1(-1+I\psi+(I-1)\psi(1-3\lambda+\lambda(2\lambda-1)\psi))-IP_1{}^2(1-\psi)}{(P_0(1+(I-1)\lambda\psi)+IP_1)^3}.$$

If λ equals one, then the derivative is

$$\frac{-P_1(1-\psi)}{(P_0(1+(I-1)\psi)+IP_1)^2}>0.$$

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

By continuity, the derivative must be negative in a region around one.

If λ equals zero, then the derivative is

$$\frac{(2I-1)\psi P_1 P_0 - I(1-\psi)P_1{}^2 - P_1 P_0}{(P_0 + IP_1)^3}$$

which is strictly positive whenever

$$(2-\frac{1}{I})\frac{\psi}{1-\psi} - \frac{1}{I(1-\psi)} > \frac{P_1}{P_0}.$$

When this condition holds, then by continuity the derivative will be positive when λ is close enough to zero.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proof of Proposition 4:

The variance of ex post beliefs is

$$\frac{P_1((I+I(I-1)\psi)+P_0+I^2P_1)}{P_0(P_0(1+(I-1)\lambda\psi)+IP_1)^2}$$

and derivative of this with respect to ψ is

$$\frac{[IP_1(1-2\lambda)+(P_0((1-2\lambda)-(I-1)\lambda\psi))](I-1)IP_1}{(P_0(1+(I-1)\lambda\psi)+IP_1)^3}.$$

This is negative if and only if

$$\lambda > (1-2\lambda)\frac{IP_1+P_0}{\psi(I-1)P_0}.$$

The variance of posterior error term is

$$\frac{P_1(I+I(I-1)\psi)+P_0(1+(I-1)\lambda\psi)^2}{(P_0(1+(I-1)\lambda\psi)+IP_1)^2}$$

and the derivative of this with respect to ψ is

$$\frac{I(I-1)P_1(IP_1+P_0-(I-1)P_0\lambda\psi(1-2\lambda))}{(P_0(1+(I-1)\lambda\psi)+IP_1)^3}.$$

This is positive if and only if

$$\frac{IP_1+P_0}{\psi(I-1)P_0} > \lambda(1-2\lambda).$$

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proof of Proposition 5:

(a) The variance of the error around the posterior when people are perfect Bayesians is

$$\frac{((1-\nu)^2+(1-\nu)\nu I+\alpha(1-\alpha)\nu^2 I^2)}{(((1-\nu)+(1-\nu+\nu I)+\alpha(1-\alpha)\nu^2 I^2)P_0+((1-\nu)I+2\alpha(1-\alpha)\nu I^2)P_1)}.$$

And the derivative of this with respect to α equals:

$$\frac{(2\alpha-1)2I^2P_1\nu(1-\nu)(1-\nu+0.5\nu I)}{(((1-\nu)^2+(1-\nu)\nu I+\alpha(1-\alpha)\nu^2 I^2)P_0+((1-\nu)I+2\alpha(1-\alpha)\nu I^2)P_1)^2}$$

Which is positive if and only if $\alpha > 0.5$. The maximum variance is

$$\frac{(1-\nu)^2+(1-\nu)\nu I}{((1-\nu)^2+(1-\nu)\nu I)P_0+(1-\nu)IP_1}$$

and the minimum variance is

$$\frac{((1-\nu)^2+(1-\nu)\nu I+0.25\nu^2 I^2)}{(((1-\nu)^2+(1-\nu)\nu I0.25\nu^2 I^2)P_0+((1-\nu)+0.5\nu I^2)}.$$

When individuals do not think that there is any common noise (i.e., they are naïve Bayesians) then the variance is

$$\frac{P_1 I(1-\nu+\nu I-2\alpha(1-\alpha)\nu I)+P_0}{(P_0+IP_1)^2}$$

and the derivative of this with respect to α equals:

$$\frac{2P_1\nu I^2(2\alpha-1)}{(P_0+IP_1)^2}$$

which is positive if and only if $\alpha > 0.5$. The maximum variance is

$$\frac{P_1 I(1-\nu+\nu I)+P_0}{(P_0+IP_1)^2}$$

and the minimum variance is

$$\frac{P_1 I(1-\nu+0.5\nu I)+P_0}{(P_0+IP_1)^2}.$$

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

The basic formula for the posterior can be written

$$\frac{P_1(\sum_{i<\alpha I}(a+b(1-\alpha))S_i+\sum_{i>\alpha I}(a+b\alpha)S_i)}{(a\,(a+b)+\alpha(1-\alpha)b^2)P_0+(a+2b\alpha(1-\alpha))IP_1}$$

where $a=1-\lambda\nu$ and $b=1-\lambda\nu I$.

With this, the variance of the difference between the posterior and D equals:

$$(a^4P_0+2a^3bP_0-2ab(b^2P_0+IP_1(2+(I-2)\nu))(\alpha-1)\alpha+b^2(\alpha-1)\alpha(IP_1(\nu-1)+ b^2P_0(\alpha-1)\alpha+2I^2P_1\nu(\alpha-1)\alpha)+a^2(b^2P_0(1-2(\alpha-1)\alpha)+IP_1(1+\nu(-1+I(1+ 2(\alpha-1)\alpha)))).$$

Divided by: $(a^2P_0+a(bP_0+IP_1)-b(bP_0+2IP_1(\alpha-1)\alpha)^2$.

The derivative of this with respect to α is then:

$$I^2P_1\nu(2\alpha-1)(IP_1(2+\nu\lambda(-6+2I^2(\nu-1)\nu(\alpha-1)\alpha\lambda^2- 2\lambda\nu(-3+\lambda\nu)-I(\lambda\nu-1)(2-(\nu+1)\lambda+4(\alpha-1)\alpha(1+(\nu-2)\lambda))))+ P_0(2+(I-2)\nu^4(1+I(\alpha-1)(I\alpha-1)\lambda^4(2\lambda-1)+\nu\lambda(-4-4\lambda+3I\lambda)+ \nu^3\lambda^3(4-12\lambda+I(-6+I(1+(I-4)(\alpha-1)\alpha(2-3\lambda)-3\lambda+15\lambda))+ \nu^2\lambda^2(12\lambda+I(3-12\lambda+I(\lambda-2(\alpha-1)\alpha(4\lambda-3))))))).$$

Divided by: $(P_0(1-\lambda\nu)^2+I(1-\lambda\nu)(P_1+P_0\nu\lambda)-I^2\nu(\alpha-1)\alpha\lambda(2P_1+P_0\nu\lambda))^3$.

When $\alpha = 0$, this becomes:

$$I^2P_1\nu(IP_1(-2+\nu\lambda(4-2\lambda\nu)+I(-2+\lambda+\nu\lambda)))+P_0(-2+\nu\lambda(2+\lambda(-3I(1+\nu)+2 (2+\nu)-(I-1)\nu(-8+I+(I-2)\nu)\lambda+2(I-2)(I-1)\nu^2\lambda^2)))).$$

Divided by: $(\lambda\nu-1)^2(P_0+IP_1+(I-1)P_0\nu\lambda)^3$.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

This is negative if and only if the other parameters satisfy:

$$IP_1(-2+\nu\lambda(4-2\lambda\nu+I(-2+\lambda+\lambda\nu)))+P_0(-2+\nu\lambda(2-\lambda(-3I(1-\nu)+2(2+\nu)-(I-1)\nu(-8+I+(I-2)\lambda)+2(I-2)(I-1)\nu^2\lambda^2)))<0.$$

And if this condition holds, then by continuity there must be some α sufficiently close to zero such this derivative is still negative.

(b) The reduction in error variance moving from $\alpha = 0$ to $\alpha = 0.5$ for the perfect Bayesian equals

$$\frac{\nu I^2 P_1[0.5+(-0.5+0.25)\nu]}{(P_0+IP_1-P_0\nu+IP_0\nu)((1+(I-2)\nu+0.25(I-2)^2\nu^2)P_0+(1+(0.5I-1)IP_1)}.$$

The reduction in error variance moving from $\alpha = 0$ to $\alpha = 0.5$ for the naïve Bayesian equals $\frac{\nu I^2 P_1}{2(P_0+IP_1)^2}$. The difference between the reduction in error variance for the naïve Bayesian and the reduction in error variance for the perfect Bayesian is:

$$\frac{1}{2}I^2P_1\nu\left(\frac{1}{(P_0+IP_1)^2}-\frac{1+(0.5I-1)\nu}{(P_0+IP_1-P_0\nu+IP_0\nu)(IP_1(1+(0.5I-1)\nu)+P_0(1+(I-2)\nu+0.25(I-2)^2\nu^2))}\right).$$

Which is greater than zero if and only if

$$\frac{1}{(P_0+IP_1)^2}>\frac{1+(0.5I-1)\nu}{(P_0+IP_1-P_0\nu+IP_0\nu)(IP_1(1+(0.5I-1)\nu)+P_0(1+(I-2)\nu+0.25(I-2)^2\nu^2)}.$$

This condition can be rewritten:

$$\nu(I-1)P_0((P_0+IP_1)+IP_1(0.5I-1)\nu+P_0((I-2)\nu+0.25(I-2)^2\nu^2))+(P_0+IP_1)((P_0((0.5I-1)\nu+0.25(I-2)^2\nu^2))>0.$$

So the naïve Bayesian always becomes more accurate.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proof of Proposition 6:

The formula for the posterior can be written:

$$\frac{P_1(D(1+\frac{I-1}{\lambda})+\varepsilon_i+\frac{1}{\lambda}\sum_{j\neq i}\varepsilon_j+\frac{(I-1)(\lambda-1)\eta}{\lambda})}{P_0+IP_1}.$$

And the difference between the posterior and the true value of the outcome, D, equals

$$-\frac{\frac{(I-1)(\lambda-1)\eta}{\lambda}P_1D-P_0D+P_1(\varepsilon_i+\frac{1}{\lambda}\sum_{j\neq i}\varepsilon_j+\frac{(I-1)(\lambda-1)\eta}{\lambda})}{P_0+IP_1}.$$

The expected value of this quantity, conditioning on η, but not on D, equals $\frac{P_1(I-1)(\lambda-1)\eta}{\lambda(P_0+IP_1)^2}$ and this is the degree of bias. This quantity is rising with λ, η, I and P_1 and falling with P_0.

The variance of the posterior equals

$$\frac{P_1^2(1+\frac{I-1}{\lambda})^2+P_0P_1(1+\frac{I-1}{\lambda^2})+P_0P_1^2(\frac{(I-1)(\lambda-1)\eta}{\lambda})^2Var(\eta)}{P_0(P_0+IP_1)^2}.$$

And the derivative of this with respect to λ equals

$$-2P_1^2\frac{I-1}{\lambda^2}(1+\frac{I-1}{\lambda})-2P_0P_1(\frac{I-1}{\lambda^3})+2P_0P_1^2\frac{(I-1)^2(\lambda-1)}{\lambda^3}Var(\eta)$$

which is positive if and only if $(I-1)(\lambda-1)Var(\eta) > \frac{1}{P_0}(I+\lambda-1)+\frac{1}{P_1}$.
The variance of the difference between the posterior belief and the true value is

$$\frac{P_1^2(\frac{(I-1)(\lambda-1)}{\lambda})^2+P_0^2+2P_0P_1(I-1)(\frac{\lambda-1}{\lambda})+P_0P_1(1+\frac{I-1}{\lambda^2})+P_0P_1^2(\frac{(I-1)(\lambda-1)}{\lambda})Var(\eta)}{P_0(P_0+IP_1)^2}.$$

Which is increasing with λ.

Within a group, the variance of beliefs equals $\frac{P_1(1+\frac{I-1}{\lambda^2})}{(P_0+IP_1)^2}$ which is declining with λ.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proof of Corollary 1:

The results in this corollary follow directly from applying Proposition 4 and that fact that $\lambda = \frac{1}{\gamma}$.

Proof of Proposition 7:

The posterior belief of the decision–maker equals $\beta^{\star}(ID + \Sigma_i \varepsilon_i + \beta^{\star}\Sigma_i \mu_i)$ and the variance of this equals $\frac{I^2\beta^{\star 2}}{P_1} + \frac{\beta^{\star 2}}{P_1} + \frac{\beta^{\star 4}}{P_\mu}$ which is increasing with $\beta^{\star}$ and hence decreasing with λ.

The error term in the posterior belief equals $(\beta^{\star}I - 1)D + \beta^{\star}(\Sigma_i \varepsilon_i + \beta^{\star}\Sigma_i \mu_i)$ which has variance $-\frac{(1-\beta^{\star}I)^2}{P_0} + \frac{\beta^{\star 2}I}{P_1} + \frac{\beta^{\star 4}I}{P_\mu}$ and the derivative of this with respect to $\beta^{\star}$ is $-\frac{2I(1-\beta^{\star}I)}{P_0} + \frac{2\beta^{\star}I}{P_1} + \frac{4\beta^{\star 3}I}{P_\mu}$. This quantity is positive as long as $\frac{\beta^{\star}I}{P_0} + \frac{\beta^{\star}}{P_1} + \frac{2\beta^{\star 3}}{P_\mu} > \frac{1}{P_0}$ which must hold when $\lambda > 1$ since $\beta^{\star}(\lambda)$ satisfies $\frac{\beta^{\star}I}{P_0} + \frac{\beta^{\star}}{P_1} + \frac{2\lambda\beta^{\star 3}}{P_\mu} = \frac{1}{P_0}$. Since the variance of the error is increasing in $\beta^{\star}$ when $\lambda < 1$, it must be decreasing with λ if this condition holds.

The perceived variance of the error term equals $\frac{(1-\beta^{\star}I)^2}{P_0} + \frac{\beta^{\star 2}I}{P_1} + \lambda\frac{\beta^{\star 4}I}{P_\mu}$ and the derivative of this with respect to λ is

$$\frac{\beta^{\star 3}I}{P_\mu}\left(\frac{4P_1P_\mu - 3\beta^{\star}IP_1P_\mu - 3\beta^{\star}IP_0P_\mu - 2\beta^{\star 2}\lambda P_0P_1}{IP_1P_\mu + P_0P_\mu + 6\lambda\beta^{\star 2}P_0P_1}\right) > 0.$$

The expected error term in the posterior equals $\beta^{\star 2}I\bar{\mu}$, and the value of $\bar{\mu}$ can be written as 0.5 times $Cov(\pi, \frac{1}{\gamma}) + (\bar{\pi})(\overline{\frac{1}{\gamma}})$ where $Cov(\pi, \frac{1}{\gamma})$ is the ex post covariance of π and $\frac{1}{\gamma}$ and $(\bar{\pi})$ and $(\overline{\frac{1}{\gamma}})$ are the ex post means of π and $\frac{1}{\gamma}$ respectively. The ex post bias is clearly increasing with $Cov(\pi, \frac{1}{\gamma})$ and $(\bar{\pi})$, and with $(\overline{\frac{1}{\gamma}})$ when $(\bar{\pi}) > 0$.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024

Proof of Proposition 8:

The posterior belief of the decision-maker equals

$$\frac{P_1(ID+\Sigma_i\varepsilon_i)}{P_0+IP_1}+\left(\frac{P_1}{P_0+IP_1}\right)^2\Sigma_i(1-\lambda_i)\mu_i.$$

If $\lambda_i=\lambda$ for all i, then the variance of this term is

$$\frac{IP_1}{P_0(P_0+IP_1)}+\left(\frac{P_1}{P_0+IP_1}\right)^4\frac{(1-\lambda)^2 I}{P_\mu}$$

which is falling with λ. The variance of the error term equals

$$\frac{1}{P_0+IP_1}+\left(\frac{P_1}{P_0+IP_1}\right)^4\frac{(1-\lambda)^2 I}{P_\mu}$$

which is also falling with λ. The perceived variance of the error term is

$$\frac{1}{P_0+IP_1}+\left(\frac{P_1}{P_0+IP_1}\right)^4\frac{(1-\lambda)^2\lambda I}{P_\mu}.$$

The derivative of this with respect to λ is

$$\left(\frac{P_1}{P_0+IP_1}\right)^4\frac{I}{P_\mu}\left[3\lambda^2-4\lambda+1\right]$$

which is negative if and only if $3\lambda^2-4\lambda+1<0$, which holds for $\frac{1}{3}<\lambda<1$.

The expected bias in decision-making equals $\left(\frac{P_1}{P_0+IP_1}\right)^2 IE((1-\lambda_i)\mu_i)$ which also equals

$$I\left(\frac{P_1}{P_0+IP_1}\right)^2(0.5E(1-\lambda_i)E(\pi_i)E(\tfrac{1}{\gamma_i})+0.5E(1-\lambda_i)Cov(\pi_i,\tfrac{1}{\gamma_i})Cov(1-\lambda_i,\mu_i)).$$

This is increasing in $Cov(\pi_i,\frac{1}{\gamma_i})$, $Cov(1-\lambda_i,\mu_i)$, and $E(\pi_i)$ and increasing with $E(\frac{1}{\gamma_i})$ and $E(1-\lambda_i)$ if and only if $E(\pi_i)$ is positive.

Downloaded from https://academic.oup.com/jla/article/1/1/263/889022 by guest on 20 March 2024