Discussion Papers
Department of Economics
University of Copenhagen

No. 20-03

Fake News, Voter Overconfidence, and the Quality of Democratic Choice

by

Melis Kartal and Jean-Robert Tyran

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Fake News, Voter Overconfidence, and the Quality of Democratic Choice*

Melis Kartal[†] **Jean-Robert Tyran**[‡]

10 June, 2020

Abstract

This paper studies, theoretically and experimentally, the effects of overconfidence and fake news on information aggregation and the quality of democratic choice in a common interest setting. We theoretically show that overconfidence exacerbates the adverse effects of widespread misinformation (i.e., fake news). We study extensions that allow for partisan biases, targeted misinformation intended to move public opinion in a specific direction, and correlated news signals (due to media ownership concentration or censure). In our experiment, voters are exposed to correct news or misinformation depending on their cognitive ability. Absent overconfidence, more cognitively able subjects are predicted to vote while less able subjects are predicted to abstain, and information is predicted to aggregate well. We provide evidence that overconfidence induces misinformed subjects to vote excessively, thereby severely undermining information aggregation.

JEL Classification: D72, D83, D91

Keywords: behavioral political economy, voting, misinformation, Dunning-Kruger effect.

*We are grateful for comments by participants of various conferences, seminars, and workshops, and would like to especially thank Amrita Dhillon, Sebastian Fehrler, Botond Köszegi, Aniol Llorente-Saguer, Akaki Mamageishvili, César Martinelli, Salvatore Nunnari, Pietro Ortoleva, Debraj Ray, and Rupert Sausgruber for their very helpful comments and suggestions.

†Vienna University of Economics and Business, Email: melis.kartal@wu.ac.at.

‡University of Vienna, University of Copenhagen, and CEPR (London), E-mail: jean-robert.tyran@univie.ac.at.

1 Introduction

Mass misinformation, now also known as *fake news*, is at the center stage of political and academic discourse because it disrupts the veracity of media coverage and may undermine public opinion.[1,2] According to a survey study by the Pew Research Center in the US, a sizable majority of respondents believes that fake news create confusion about basic facts. However, 84% of these respondents are (somewhat or very) confident in their ability to recognize fake news. According to another survey study (Jang and Lim, 2018), individuals believe that fake news have a greater effect on others than themselves. These studies suggest overconfidence in one's ability to sort fact from fiction and that voters may be more prone to vote on the basis of misinformation than they think.

In this paper, we study theoretically and experimentally the joint effect of overconfidence and the dissemination of fake news on public opinion and the quality of democratic choice. We consider the classic common-interest voting environment with two policy alternatives where citizens share the objective to select the better policy but may differ in their information and hence disagree about which policy is better.[3] Abstention is allowed as in Feddersen and Pesendorfer (1996). In contrast to the previous literature, individuals may be overconfident (or underconfident) in their *competence* to obtain accurate news and form correct opinions. We theoretically analyze the effects of such biased perceptions on voting behavior and information aggregation, and we also corroborate some theoretical predictions of our model in the laboratory. Our model shows not only does overconfidence undermine information aggregation, but perhaps more importantly it can gravely exacerbate the impact of pervasive misinformation on democratic decision making. Our framework also relates to the rationale behind the supply of misinformation as it shows how the interaction between

[1] We define *misinformation* and *fake news* as fabricated news usually with a deliberate intention to deceive (see also Allcott and Gentzkow (2017) and Lazer *et al.* (2018)). Thus, misinformation and fake news represent a significant subset of the broad category of *false news*, which may be deliberately or unintentionally misleading.

[2] While political fake news stories and the media coverage of the fake news phenomenon have gained an unprecedented level of prominence recently, the concern that the mass media often involves misinformation and false news goes back to Lippmann and Merz (1920), and Lippmann (1922). See also Hermann and Chomsky (2003).

[3] We interpret voting outcome as public opinion since (i) public opinion consists of individual opinions; (ii) individual opinions are reflected in political activity; and (iii) we interpret voting as political activity (whether it be through the ballot, contacting political representatives, lobbying, participation in a demonstration, or providing an opinion in a poll instead of saying "No opinion") and abstention the lack thereof.

overconfidence and misinformation can benefit a third party or partisan voters (when we relax the common-interest assumption).

The electorate decides on a policy by majority vote. Individuals vote for one of two policies, a and b or abstain. Individual payoffs depend on the voting outcome and the underlying state of the world. There are two states of the world, A and B. The ex-ante probability that the state is A is common knowledge. In the baseline model, individuals have identical preferences and prefer the chosen policy to match the state. In other words, they strictly prefer (i) policy a to be chosen if the state is A, and (ii) policy b to be chosen if the state is B. Individuals do not know the state of the world. Instead, each individual privately observes a binary news signal. The precision of signals is heterogeneous as individuals differ in their competence. Therefore, as in Feddersen and Pesendorfer (1996) and the pursuant literature, individuals who are less confident in their signal may rationally refrain from voting for their signal.

Less competent individuals are more vulnerable to misinformation in our model. Moreover, as discussed above, individuals hold subjective (and possibly inflated) views about their competence (e.g., the quality of the news they receive and the accuracy of their opinions). Hence, the rational mechanism in Feddersen and Pesendorfer in which less competent individuals refrain from following their signal (or, similarly, a rational mechanism which leads such individuals to acquire more information in an extended model) may not work with the *Dunning-Kruger effect.* According to the Dunning-Kruger effect, incompetent individuals prone to overconfidence do not recognize their lack of competence. That is, they are "unskilled and unaware of it." Thus, our hypothesis is that incompetent individuals are (i) more likely to be exposed to fake news and less able to discern correct and fake news, but (ii) many of them are unaware and assess their ability as substantially greater than it is. As a result, they are more likely to act based on misinformation.[4]

We theoretically show that information aggregation is undermined in the presence of Dunning-Kruger effect. Moreover, overconfidence tends to be more troubling as the veracity of media coverage declines (e.g., media veracity may decline due to misinformation becoming more prevalent). To illustrate, consider a simple example with three individuals, where each individual is equally likely to be competent or incompetent. Each individual privately

[4]In a more realistic setup, recognizing that one has low competence on an issue may lead to information acquisition rather abstention. However, in either case, the absence of overconfidence is essential, and therefore, we begin our analysis in a simpler setup and leave the addition of costly information acquisition to the model to future research (see also Footnote 5).

observes their type, and this observation is either correct or overconfident. Assume that every competent individual obtains a perfectly informative signal and learns the true state, and every incompetent individual obtains a noisy signal which matches the true state with probability 0.7—this is a setting with high media veracity. The effect of overconfidence is trivial in such a setting because the average signal precision is quite high at 0.85. Next, assume that media veracity decreases making it more difficult for both types of individuals to infer the state correctly: a competent individual obtains a signal which matches the state with probability 0.9, and an incompetent individual obtains an uninformative signal (i.e., the signal matches the state with probability 0.5). Comparing these two scenarios, the effect of overconfidence is much more pronounced when the media veracity is low.[5] This is also why our experiment focuses on a setting with a high share of misinformation.

As this example suggests, overconfidence is typically more harmful at low levels of media veracity acting as a "multiplier" of misinformation dissemination. Hence, the powerful impact of the fake news phenomenon relies on the Dunning-Kruger effect in our model—an abundance of unskilled and unaware voters. More generally, the quality of public opinion and democratic decision making depends on both the prevalence of misinformation in circulation and the extent of the Dunning-Kruger effect in the population.

We also analyze the implications of richer models with partisan voters, differing media veracities in different states of the world, and correlated news signals. The latter two models are particularly relevant given the extent of media ownership concentration. Differing media veracities in different states of the world may be caused by a third party (e.g., a special interest, corporate giant, or government) that strictly prefers a specific policy regardless of the state of the world and disseminates self-serving "news." Dissemination of such news gives rise to an asymmetry in the veracity of media coverage in different states of the world as those news are benign in one state, whereas in the other state they are misleading.[6] We show that the special interest can benefit from those news but *only* in the presence of a high degree of overconfidence. Also, we are the first to analyze a model of information aggregation with correlated signals (that is, only a few or several media outlets generate

[5] The same conclusion also holds if incompetent voters can acquire information (i.e., become competent) at a positive but not too high cost.

[6] In addition to media concentration and implicit or explicit media censorship, technological progress seems to have provided third parties and special interests with novel tools to sway public opinion. For example, findings of Marlow, Miller, and Roberts (2020) suggest that online bots have a substantial role in amplifying denialist messages about climate change as well as support for Trump's withdrawal from the Paris Agreement. See also our discussion in Section 2.3.

independent news signals, and multiple individuals receive the exact same news signal). The ubiquitous assumption of the information aggregation literature that every individual observes an independent signal is hardly realistic. According to Bagdikian (2004), there is substantial concentration in media ownership — for example, the number of corporations controlling most of the media in the US decreased to 5 from around 50 in 1983. We identify conditions under which overconfidence and correlation of news signals jointly undermine information aggregation. Note that increased media concentration can also facilitate the third-party interference with media mentioned above.

To illustrate the impact of overconfidence on information aggregation, we report the results of an experiment. In light of our discussion above, the experimental design involves a setting with a high degree of misinformation and is conducive to overconfidence. We implemented subjective beliefs regarding signal precisions as follows. Subjects take an incentivized quiz with math and logic puzzles before (they learn about) the voting stage. Subjects are not informed about their score in the quiz until the end of the experiment. During the voting phase, participants are explained that the signal they get regarding the state depends on their quiz score; that is, a participant observes a signal that has the same color as the true state (i.e., correct news) if and only if the participant's score places him in the top 1/3 of all the subjects in the same experimental session. Since subjects do not learn whether or not they are in the top 1/3 until the end of the experiment, they must form a belief regarding their likelihood of being in top 1/3 (i.e., they must form a belief about the informativeness of their signal). We elicit subjects' beliefs regarding their relative score in an incentivized manner. Our experimental findings reveal substantial overconfidence, which then translates into excessive turnout and very poor results in group decision-making consistent with our theoretical prediction. In fact, we find that the inferior policy is chosen all the time. We rule out explanations other than overconfidence for excessive turnout.[7] These results indicate that collective overconfidence can result in more extreme outcomes than individual overconfidence—while collective decision making can theoretically cancel out modest levels of individual overconfidence, at high levels (as in our experiment) it can result in drastic inefficiency.[8]

[7] For example, in a control phase in which subjects learn the objectively correct precision of their signals as in the standard information aggregation experiments, we observe a moderate turnout rate and highly successful aggregation of information.

[8] Esponda and Vespa (2015) has shown that experimental subjects have difficulty in extracting information from hypothetical events such as pivotality consistent with previous theoretical work of Eyster and

Our theory and experiment suggest that a pervasive Dunning-Kruger effect severely undermines information aggregation and democratic decision making provided that a sufficiently high fraction of news is false. This is in line with the idea that the resilience of democratic decision making relies on other institutions. To uphold the democratic ideal, it is clearly essential to make high quality education widely accessible to the public to limit the fraction of unskilled and unaware voters who are more vulnerable to manipulative news.[9] Furthermore, news media has been dubbed the "fourth estate" because it has an essential role in maintaining checks and balances and in limiting the power of special interests. For the media to fulfill this role, there must be numerous *high-quality* news sources investigating and reporting independent of commercial or political interests, which limits on media ownership concentration and extensive public support for media (e.g., as in Finland and Germany) may make easier to achieve.

2 Model

An electorate of N citizens must choose one of two policies: policy a and policy b. Each individual casts a vote for one of the two policies or abstains. The policy that receives a majority of the votes is the chosen policy, and ties are broken randomly. The utility of each individual depends on the chosen policy as well as the realization of a stochastic state of the world $S \in \{A, B\}$. Citizens have identical preferences and agree that policy a is superior in state A and policy b is superior in state B. We normalize citizens' utilities without loss of generality and assume that

$$\begin{aligned} u(a|A) &= u(b|B) = 1 \\ u(a|B) &= u(b|A) = 0. \end{aligned} \tag{1}$$

Citizens do not know the state of the world, and the common prior is that the state is A with probability $\pi \in (0, 1)$. Moreover, each individual i privately observes a noisy news signal $s_i \in \{\alpha, \beta\}$ regarding the state of the world. The term q_i represents the precision of individual i's signal, and $q_i = \Pr(s_i = \alpha|S = A) = \Pr(s_i = \beta|S = B)$. Thus, the higher the

Rabin (2005) and Esponda (2008). However, in our experiment subjects' possible failure to understand the pivotality logic is largely irrelevant.

[9] Martin Luther King (1947) writes that protecting individuals from propaganda is "one of the chief aims of education. Education must enable one to sift and weigh evidence, to discern the true from the false, the real from the unreal, and the facts from the fiction."

magnitude of q_i, the higher the probability that i observes correct news (i.e., a signal that matches the true state).

For every i, q_i is an independent draw from a common distribution denoted by F. The distribution F has support $[\underline{q}, \bar{q}]$, where $\underline{q} < \bar{q}$ and $\bar{q} > 0.5$, and F has density f, which is positive on all of its support. The distribution F is common knowledge, whereas q_i is i's private information. Our underlying hypothesis is that individuals differ in their competence and ability to discern the true state (for example because individuals are heterogenous in reasoning skills and the ability to tell apart proper news sources from dubious sources). Thus, we interpret q_i as a measure of skill and competence — fixing the veracity of media coverage. More generally, signal precision q_i depends not only on competence but also on media veracity, which is exogenously fixed for the time being.[10] We do not explicitly model media veracity in this model; however, $F[\underline{q}, \bar{q}]$ and $\int_{\underline{q}}^{\bar{q}} q dF$ serve as indicators of media veracity.

The relevant characteristic of individual i is not only q_i, but also a possible *bias*: i is either biased or unbiased in their perception of q_i.[11] More formally, i perceives q_i as $p_i(q_i) \in [\underline{q}, \bar{q}]$, and $p_i(q_i) = q_i$ (that is, the perception of i is correct) if i is unbiased, and $p_i(q_i) \neq q_i$ if i is biased. A biased individual is either overconfident or underconfident. If i is overconfident, then $p_i(q_i) > q_i$ (that is, i's perception of q_i is inflated), whereas if i is underconfident, $p_i(q_i) < q_i$. We assume without loss of generality that $p_i(q) \equiv p_o(q)$ for every overconfident i, and $p_i(q) \equiv p_u(q)$ for every underconfident i, where $p_o(q)$ and $p_u(q)$ are continuous and increasing in q. Thus, for every i and q, $p_i(q)$ takes one of three possible values: $p_o(q)$, q, and $p_u(q)$.[12]

An individual with signal accuracy q is overconfident with probability $\lambda_o(q) > 0$, underconfident with probability $\lambda_u(q) \geq 0$, and unbiased with the remaining probability,

[10] It is possible to model precision q_i as a function of competence of i, which is an independent draw from a common distribution of competence and media veracity, which is the same for every individual. As a result, the rate of information aggregation in the society depends on both the competence distribution in the society and the media veracity. We will revisit this point in Section 2.3 when we discuss extensions.

[11] Many individuals may derive an "ego utility" from positive views about their skills and competence (Kőszegi, 2006). Therefore, overconfidence in q_i could be explained by ego utility derived from inflated views about one's competence in the domain of politics and policy issues.

[12] Results extend to settings with various levels of overconfidence and underconfidence, for example, settings in which there exist finitely many overconfidence functions $p_o^j(q) \in (q, \bar{q}]$ and probabilities $\lambda_o^j(q)$ with $j \in \{1, ..., J\}$ such that $p_o^j(q) < p_o^{j+1}(q)$ for every q and $j \in \{1, ..., J-1\}$ and $\lambda_o^j(q)$ is the respective probability that an individual with skill q perceives it as $p_o^j(q)$. See the proof of Proposition 1 in Online Appendix A.1.

where $\lambda_o(q)$ and $\lambda_u(q)$ are continuous. While we do not need to impose more structure on $\lambda_o(q)$ or $\lambda_u(q)$ until Section 2.3, the most relevant scenario is the case where $\lambda_o(q)$ is high at low levels of q. In this scenario, overconfidence is particularly prevalent at low competence levels representing the Dunning-Kruger effect.

2.1 Equilibrium Analysis

The strategy for individual i, denoted by σ_i, maps subjective signal accuracy $p_i(q_i) \in \{p_o(q_i), q_i, p_u(q_i)\}$ and signal s_i to voting for a, voting for b or abstention; i.e., $\sigma_i : [\underline{q}, \bar{q}] \times \{\alpha, \beta\} \to \{a, b, 0\}$, where 0 represents the decision to abstain.[13] A profile of strategies $\{\sigma_i\}_{i \leq N}$ constitutes a Bayesian Nash equilibrium if σ_i is a best response to others' strategies σ_{-i} for every i.

Every i believes that their perception of q_i is unbiased. As for the belief regarding others, we assume in the main text that individuals are *unaware* of perception biases unless otherwise stated. Put differently, every i believes that every j perceives q_j correctly (i.e., $p_j(q_j) = q_j$) and acts on the basis of correct beliefs as in the standard setting. All of our main results in this section, and many of our results in the upcoming sections are robust to awareness of others' perception biases (see also Online Appendix A.2).

For tractability, we focus on symmetric Bayesian Nash Equilibria in which all citizens who receive the same signal choose the same strategy; that is, we look for a voting strategy of the form $\sigma : [\underline{q}, \bar{q}] \times \{\alpha, \beta\} \to \{a, b, 0\}$, which is the same for all individuals. We rule out unresponsive equilibria in the analysis of equilibrium behavior. Such equilibria are straightforward as strategies do not rely on s_i. Hereafter, Bayesian Nash equilibrium refers to symmetric and responsive Bayesian Nash Equilibrium.[14]

We now show that every Bayesian Nash equilibrium consists of "cutoffs" (with and without perception biases). To demonstrate how equilibrium cutoffs arise, we start by assuming that the signal of individual i is α. In that case, i weakly prefers voting for a over abstention if and only if

$$\frac{1}{2}\left(\Pr(piv_a \cap S = A | s_i = \alpha) - \Pr(piv_a \cap S = B | s_i = \alpha)\right) \geq 0,$$

where piv_a denotes the event where i's vote for policy a is pivotal fixing others' strategies

[13] Limiting σ_i to pure strategies is for notational convenience. We allow for randomization in the proofs.

[14] In a responsive equilibrium, i's strategy is not independent of s_i.

$\sigma_{-i} \equiv \sigma$. It can be checked that the inequality above can be written as

$$p_i(q_i) \geq \frac{(1-\pi)\Pr(piv_a|S=B)}{\pi \Pr(piv_a|S=A) + (1-\pi)\Pr(piv_a|S=B)} = \Pr(S=B|piv_a)$$

by Bayes rule and conditional independence (e.g., $\Pr(s_i = \alpha \cap piv_a|S = A) = \Pr(s_i = \alpha|S = A)\Pr(piv_a|S = A)$). Hence, in order to choose between voting for a and abstaining, individual i compares the precision of $s_i = \alpha$ to $\Pr(S = B|piv_a)$. Note that $\Pr(S = B|piv_a)$ depends not only on σ_{-i} but also on the presence of perception biases and whether or not individuals are aware of others' possible biases.

Next, consider the preference of i for voting for policy a over voting for policy b after receiving an α signal. The respective expected utility from voting for a and b can be written as

$$\frac{1}{2}\left(\Pr(piv_a \cap S = A|s_i = \alpha) - \Pr(piv_a \cap S = B|s_i = \alpha)\right)$$

and

$$\frac{1}{2}\left(\Pr(piv_b \cap S = B|s_i = \alpha) - \Pr(piv_b \cap S = A|s_i = \alpha)\right).$$

Thus, i prefers voting for a over voting for b if the latter term is higher than the former term. By Bayes rule and conditional independence, it follows that

$$p_i(q_i) \geq \frac{(1-\pi)\left(\Pr(piv_a|S=B) + \Pr(piv_b|S=B)\right)}{\pi(\Pr(piv_a|S=A) + \Pr(piv_b|S=A)) + (1-\pi)(\Pr(piv_a|S=B) + \Pr(piv_b|S=B))}$$

must hold, which gives

$$p_i(q_i) \geq \frac{\Pr(S=B|piv_a)\Pr(piv_a) + \Pr(S=B|piv_b)\Pr(piv_b)}{\Pr(piv_a) + \Pr(piv_b)}.$$

Combining the results above, individual i with $s_i = \alpha$ weakly prefers voting for a over other options if and only if

$$p_i(q_i) \geq \max\left\{\Pr(S=B|piv_a), \eta \Pr(S=B|piv_a) + (1-\eta)\Pr(S=B|piv_b)\right\},$$

where $\eta = \frac{\Pr(piv_a)}{\Pr(piv_a)+\Pr(piv_b)}$. Moreover, individual i with $s_i = \alpha$ weakly prefers voting for b over other options if and only if

$$p_i(q_i) \leq \min\left\{\Pr(S=B|piv_b), \eta \Pr(S=B|piv_a) + (1-\eta)\Pr(S=B|piv_b)\right\}$$

Similarly, it can be shown that if i has a β signal, i weakly prefers voting for b if and only if

$$p_i(q_i) \geq \max\left\{\Pr(S=A|piv_b), (1-\eta)\Pr(S=A|piv_b) + \eta \Pr(S=A|piv_a)\right\}$$

and voting for a if and only if

$$p_i(q_i) \leq \min\{\Pr(S=A|piv_a), (1-\eta)\Pr(S=A|piv_b) + \eta\Pr(S=A|piv_a)\}.$$

Lemma 1 characterizes equilibrium voting behavior and formally states that it has a cutoff structure based on the analysis above.[15]

Lemma 1 *Every Bayesian Nash equilibrium consists of two cutoffs q^a and q^b such that* (1) *an individual votes for a if and only if either i's signal is α and $p_i(q_i) \geq q^a$ or i's signal is β and $p_i(q_i) \leq 1-q^a$; and* (2) *an individual votes for b if and only if either i's signal is β and $p_i(q_i) \geq q^b$ or i's signal is α and $p_i(q_i) \leq 1-q^b$, where*

$$q^a = \max\{\Pr(S=B|piv_a), \eta\Pr(S=B|piv_a) + (1-\eta)\Pr(S=B|piv_b)\} \tag{2}$$

and

$$q^b = \max\{\Pr(S=A|piv_b), \eta\Pr(S=A|piv_a) + (1-\eta)\Pr(S=A|piv_b)\}. \tag{3}$$

In particular, $q^a = \Pr(S=B|piv_a)$ and $q^b = \Pr(S=A|piv_b)$ if the correct policy is chosen with a probability greater than 0.5 in both states. If $\pi = 0.5$, then $q^a = \Pr(S=B|piv_a)$ and $q^b = \Pr(S=A|piv_b)$ always hold, and if in addition $\underline{q} = 0.5$, then $q^a = q^b$.[16]

We now investigate the effect of perception biases on collective decision making. It is not possible to show the uniqueness of the responsive equilibrium characterized in Lemma 1. Therefore, our analysis focuses hereafter on the optimal "unbiased" equilibrium as the comparison benchmark.[17] We also take into account unresponsive equilibria from now on because the optimal equilibrium is an unresponsive one if π is sufficiently close to 0 or 1.

In Lemma 2, we characterize sufficient conditions under which the optimal equilibrium involves at least one interior cutoff—that is, at least one of the cutoffs q^a, q^b, $1-q^a$, and

[15] While perception biases and whether or not individuals are aware of the presence of perception biases do not have an impact on the general characterization of the equilibrium as outlined in Lemma 1, the precise equilibrium depends on whether or not there are perception biases, and whether or not individuals are aware of the presence of biases. For instance, the equilibrium of an unbiased electorate is *always* an equilibrium with perception biases and unawareness regarding biases, but this is not necessarily true with awareness.

[16] Note that $p_i(q_i) \geq q^a$ and $1-p_i(q_i) \geq q^b$ can never be satisfied at the same time by the same $p_i(q_i)$ (except for the indifference case, where $p_i(q_i) = q^a = 1-q^a$) because $q^a + q^b \geq 1$ must always hold in equilibrium given the definitions of q^a and q^b).

[17] The optimal equilibrium is simply the one that maximizes the prior-weighted average probability that the correct policy is chosen. In case there are multiple optimal equilibria in the unbiased case for fixed π, this is inconsequential for the upcoming results.

$1 - q^b$ is an element of $(\underline{q}, \bar{q})$.[18] To see why an interior cutoff matters regarding the impact of perception biases, consider the following example. If $\pi = \underline{q} = 0.5$ with N being an odd number, the case in which every i votes for the policy that matches s_i (regardless of q_i) is an equilibrium. Thus, in this equilibrium $q^a = q^b = \underline{q}$, and there is no interior cutoff. If this equilibrium turns out to be the optimal equilibrium in the unbiased case, then overconfidence does not have an impact on efficient decision making because, without an interior equilibrium cutoff, it has no effect on equilibrium behavior.[19] Nevertheless, Lemma 2 shows that the case where $\underline{q} = \pi = 0.5$ can be considered a knife-edge situation. Put differently, for every $\underline{q} \geq 0$, there exists a nontrivial set of π values such that the optimal equilibrium involves at least one interior cutoff in an unbiased electorate.[20]

Lemma 2 *Assume that the electorate is unbiased.* (i) *If $\underline{q} < 0.5$, there exists a $\pi^* \in (0.5, 1)$ such that for all $\pi \in (1-\pi^*, \pi^*)$ the optimal Bayesian Nash equilibrium has an interior cutoff.* (ii) *If $\underline{q} = 0.5$ and N is even, there exists a $\pi^* \in (0.5, 1)$ such that for all $\pi \in (1 - \pi^*, \pi^*)$ the optimal Bayesian Nash equilibrium has an interior cutoff.* (iii) *If $\underline{q} = 0.5$ and N is odd or if $\underline{q} > 0.5$, then there exists a $\pi^* \in (\underline{q}, 1)$ such that for all $\pi \in (1 - \pi^*, 1 - \underline{q}) \cup (\underline{q}, \pi^*)$ the optimal Bayesian Nash equilibrium has an interior cutoff.*

Proposition 1 shows that under the parameters specified in Lemma 2 deviations from the optimal equilibrium strategy (e.g., due to overconfidence) are harmful. The intuition is as follows. In common interest voting games, the optimal symmetric strategy—if it exists—is an equilibrium strategy as shown by McLennan (1998). Put differently, what is not a best response for an individual cannot be optimal for the group as a whole. Under the conditions spelled out in Lemma 2, the equilibrium behavior of biased individuals is not a best response as biased individuals deviate from the presumed best-response strategy—for example, if $q^a \in (\underline{q}, \bar{q})$, then the actual cutoff overconfident individuals implement is $p_o^{-1}(q^a) < q^a$. As a result, information aggregates at a lower rate in the presence of perception biases (relative to the optimal unbiased equilibrium). While Proposition 1 substantially owes to the insight in McLennan (1998), the qualification is that showing the existence of the optimal strategy and

[18] We prove the existence of a responsive equilibrium under the conditions stated in Lemma 2 in Corollary 1 in Online Appendix A.1.

[19] We mostly consider the impact of overconfidence in the main text. Underconfidence does not change our main results and can only make the impact of overconfidence worse. For example, highly competent individuals who are sufficiently underconfident will not vote for their signal whereas they should in order to improve information aggregation.

[20] We explicitly characterize (a subset of) these π values in the proof of Lemma 2.

characterizing when overconfidence results in a true deviation from it is extremely difficult in our setup.

Proposition 1 *Under the conditions on $\underline{q}$ and π stated in Lemma 2, the decision making accuracy of the optimal Bayesian Nash Equilibrium of an unbiased electorate cannot be attained in an electorate with biased citizens.*

Thus, overconfidence may have a harmful impact on information aggregation even if $\underline{q} > 0.5$, but comparing the condition on π for $\underline{q} > 0.5$ with that for $\underline{q} < 0.5$ in Lemma 2 suggests that the negative effect of overconfidence is likely limited if $\underline{q} > 0.5$. In fact, this effect will certainly vanish in large elections as we discuss in the next section. Therefore, we are especially interested in the case where $\underline{q} < 0.5$, which we associate with the widespread presence of misinformation and fake news (we also implement this case in our experimental study). Empirically, $q < 0.5$ may hold for a large fraction of people: a recent analysis of 126,000 stories spread in social media found that fake news reached more people than the truth (Vosoughi, Roy, and Aral, 2018). In a model with rational and unbiased individuals, observing a signal with $q < 0.5$ is nothing special. However, overconfidence may prevent individuals them from perceiving the misleading nature of their news signal. Put differently, overconfident individuals who are also susceptible to fake news may easily take misleading or fictitious news at their face value and act upon them. Thus, the impact of overconfidence is much more likely to be pronounced when many individuals are more likely to receive misinformation than correct information.

2.2 An alternative interpretation and model

In a slightly different formulation of the model with $\underline{q} < 0.5$, we can interpret $F[\underline{q}, \bar{q}]$ as the distribution of the accuracy of individuals opinions. Then, the case where $q_i < 0.5$ is directly linked to not only misinformation but also the presence of overconfidence bias with $p(q_i) \geq 0.5$, as otherwise an individual cannot rationally hold on to a belief or opinion that they objectively understand is more likely to be false than correct. In this model with $\underline{q} < 0.5$, not only does $p_i(q_i)$ differ from q_i for some individuals, but also the distribution of q_i differs between the biased electorate and its unbiased version (in an unbiased electorate, q_i must not fall below 0.5). As a result, this is a more intricate model with two different q_i distributions. However, equilibrium characterization in Lemma 1 and other main results hold in this type of model under mild assumptions. For example, one possible assumption is that the distribution of q_i is identical in the biased electorate and its unbiased version with the exception that $q_i = 0.5$ in the unbiased case for every i in the biased electorate such that $q_i < 0.5$.

2.3 Large Elections and Extensions

While π^* spelled out in Lemma 2 goes to one as N goes to infinity, and thus, the negative impact of overconfidence extends to almost all π values in large elections, this negative impact also becomes vanishingly small provided that $\underline{q} \geq 0.5$. More generally, information will aggregate (that is, the electorate will make the correct decision in both states with a probability that goes to one as N goes to infinity) despite perception biases provided that $\underline{q} \geq 0.5$ because in our framework so far,

(i) citizens have identical preferences (e.g., there are no partisans),
(ii) veracity of media coverage is identical in both states (i.e., $\Pr(s_i = \alpha|S = A) = \Pr(s_i = \beta|S = B) \geq 0.5$ if $\underline{q} \geq 0.5$),
(iii) realizations of signals (s_i) are conditionally independent, and
(iv) the voting rule is majority.

In fact, the limiting impact of overconfidence on information aggregation as N goes to infinity depends not only on $\underline{q}$ but also on $\mathbb{E}(q)$ and the extent of overconfidence in the population. In particular, the share of misinformation in news circulation must be sufficiently high so that $\mathbb{E}(q)$ is further below 0.5, and the Dunning-Kruger effect must be sufficiently pervasive for the negative impact of overconfidence to be bounded away from zero in large elections—this forms the basis of our experimental design.[21] These conditions are admittedly very strong. However, as mentioned above misinformation may indeed have a wider reach than correct information (Vosoughi, Roy, and Aral, 2018). Moreover, if these conditions hold, the result is severe: the incorrect policy is certainly chosen in the limit in at least one state of the world (see Online Appendix A.4 for the formal result and proof). This relates to the concern in many countries over reported meddling by outsiders using misinformation. For instance, in 2017 the then Prime Minister of UK Theresa May has publicly accused Russia of "planting fake stories" in order to "sow discord in the West" (*The Economist*, November 23, 2017). According to an analysis by the European Commission, Russian groups carried out a widespread misinformation campaign aimed at influencing the European Parliament election in May 2019 (*Politico*, June 14, 2019). A report prepared for the European Parliament

[21] As mentioned before, underconfidence can only make the impact of overconfidence worse.

during the COVID-19 crisis alleged that Russia and China are "driving parallel information campaigns, conveying the overall message that democratic state actors are failing and that European citizens cannot trust their health systems." (*Nature*, May 27, 2020)

Before relaxing conditions (i)-(iii) mentioned above, note that we interpret the voting rule in (iv) as the mapping of public opinion to policy outcomes, which then determines social welfare. Depending on the issue, even a minority public opinion may have a nontrivial impact on social welfare if for example the policy implementation cannot be stringent for various reasons (or requires a supermajority). One example is the growing anti-vaccination movement and public health problems this has caused. Some diseases, such as measles, require 95 percent of the population to be vaccinated in order to achieve the so-called herd immunity, and only very recently, some countries started imposing penalties for unvaccinated children. This situation can be interpreted as a mapping of public opinion into policy outcomes that depends on the *margin* of majority (unlike the majority rule). We conjecture that sometimes the limiting effect of overconfidence will be strictly negative when the margin of majority matters, while it is zero under simple majority rule. However, a formal analysis of different rules with perception biases is beyond the scope of this text.[22]

We now relax conditions (i)-(iii) and analyze richer variations of our model to characterize conditions under which overconfidence has a strictly negative impact on information aggregation in the limit as N goes to infinity. We impose no assumption on $\underline{q}$ other than $\underline{q} < \bar{q}$ unless otherwise stated. Note that characterization of equilibrium behavior in Lemma 1 extends to first two variations with minor modifications.

(i) In this model, every individual's utility function involves not only a common interest component but also an idiosyncratic preference bias. That is, the utility function of every individual is given by

$$u(a|A) = 1 = u(b|B) = 1 \geq u(a|B) = \gamma \geq u(b|A) = 0$$

with probability p_a, by

$$u(a|A) = 1 = u(b|B) = 1 \geq u(b|A) = \gamma \geq u(a|B) = 0$$

[22] More formally, the margin of majority matters under a supermajority rule or a rule in which the policy outcome is a weighted average of policies a and b with the weights being increasing functions of the vote shares for a and b as in Kartal (2015), and Herrera et al. (2019a, 2019b). Herrera et al. (2019a, 2019b) analyze information aggregation under different rules in a Poisson voting framework but without perception biases.

with probability p_b, and by (1) with the remaining probability (i.e., $\gamma = 0$). For every individual who has a preference bias for a or b, γ is an i.i.d. draw from a common distribution G with support $\Gamma \subseteq [0,1]$ (an asymmetry in such preferences can then be generated by the difference between p_a and p_b). We assume that F and G are independent.

Equilibrium characterization is analogous to that in Lemma 1 but involves a rational *motivated reasoning* mechanism (Kunda, 1990; Taber and Lodge, 2006). For brevity in notation we assume that $p(q_i) = q_i$. Let q^a and q^b be as defined in (2) and (3) respectively. Equilibrium strategies are as follows.

- Individual i with preference bias for policy a votes for a if and only if either $s_i = \alpha$ and $\frac{q_i}{q_i+(1-\gamma)(1-q_i)} \geq q^a$ or $s_i = \beta$ and $\frac{1-q_i}{(1-\gamma)q_i+1-q_i} \geq q^a$. Also, i votes for b if and only if either $s_i = \beta$ and $\frac{(1-\gamma)q_i}{(1-\gamma)q_i+1-q_i} \geq q^b$ or i's signal is $s_i = \alpha$ and $\frac{(1-\gamma)(1-q_i)}{q_i+(1-\gamma)(1-q_i)} \geq q^b$.
- Individual i with preference bias for policy b votes for a if and only if either $s_i = \alpha$ and $\frac{(1-\gamma)q_i}{(1-\gamma)q_i+1-q_i} \geq q^a$ or $s_i = \beta$ and $\frac{(1-\gamma)(1-q_i)}{q_i+(1-\gamma)(1-q_i)} \geq q^a$. Also, i votes for b if and only if either $s_i = \beta$ and $\frac{q_i}{q_i+(1-\gamma)(1-q_i)} \geq q^b$ or $s_i = \alpha$ and $\frac{1-q_i}{(1-\gamma)q_i+1-q_i} \geq q^b$.

Thus, compared to the equilibrium strategy in the baseline model with $\gamma = 0$, the magnitude of q_i is diluted or strengthened depending on the realization of s_i to the extent that the individual has a preference bias for a or b (hence, the rational motivated reasoning).

We now focus on aggregation of information. Example 1 below shows how information aggregation may be harmed by the joint effect of preference biases and the Dunning-Kruger effect.

Example 1: Assume that $\pi = 0.6$ and $p_a = 0 < p_b$ (that is, there is no preference bias for policy a). There are three possible γ-types: $\gamma_L = 0$, $\gamma_M = 0.5$, and $\gamma_H = 1$ with cumulative probabilities $G(0) = \frac{2}{5}$, $G(0.5) = \frac{3}{5}$, and $G(1) = 1$. We assume that all γ_H-type individuals vote for policy b regardless of their signal (e.g., they are partisans). There are two types of individuals in terms of competence: low type individuals with $q_i = 0.5$ and high type individuals with $q_i = 1$ where $F(0.5) = \frac{2}{3}$. In this setting, any equilibrium in which the correct policy is chosen in both states with a probability that goes to one as N goes to infinity requires many low type individuals with $\gamma_i \in \{\gamma_L, \gamma_M\}$ to vote for policy a due to the presence of partisans. If however sufficiently many of those individuals (for example, 50 percent of them) are overconfident so that $p_i(q_i) = 1$, then they vote for the policy that matches their signal and render information aggregation in state A impossible. This is because the expected relative turnout rate for policy a is lower than 50 percent, and as a result, the probability that policy a is chosen in state A goes to zero as N goes to infinity. This holds regardless of whether or not individuals are aware of others' biases. In this example, the failure of information aggregation in state A stems from the joint effect of preference biases and overconfidence: in the absence of partisans, information will aggregate even if all low type individuals are overconfident and vote for their signal.

We now provide a more formal result concerning information aggregation with preference biases and overconfidence. For tractability, we simplify the setting above assuming that $u(a|B) = 1$ with probability p_a, $u(b|A) = 1$ with probability p_b, and $u(a|B) = u(b|A) = 0$ with the remaining probability, where $p_b \geq p_a$ without loss of generality. We also assume that $p_b < 0.5$ so that information aggregation is essential for efficiency.[23]

Proposition 2 "A partisan voter group may benefit from low media veracity and high voter overconfidence" *Assume that π is bounded away from 0 and 1, and $p_a \leq p_b < 0.5$. Then, an unbiased electorate always makes the correct decision with a probability that goes to one in both states as N grows without bound. Next, let media veracity be such that*

$$p_b > p_a + (1 - p_a - p_b)\left(\int_{\underline{q}}^{\bar{q}} q dF - \int_{\underline{q}}^{\bar{q}} (1-q) dF\right) > p_a.$$

If $\bar{q} = 1$, then a highly overconfident electorate will choose policy b with a probability that goes to one in both states. If $\bar{q} < 1$, the same result obtains if in addition $\bar{q} \geq 1 - \underline{q}$ and $p_a + (1 - p_a - p_b)\int_{0.5}^{\bar{q}} (1-q) dF > p_b$ holds.

(ii) In this model, we relax the assumption that the distribution of signal precisions is identical in the two states. That is, the distribution of q_i may depend on the state of the world, and thus it is also subject to aggregate uncertainty. This is in fact a special case of the setting where not only the state of the world but also the state of media veracity is stochastic determining the distribution of signal precisions in the two states. The two types of states may or may not be independent. Example 2 below assumes that the two types of states are independent and shows that even if every individual is *on average* more likely to be correctly informed than misinformed in both states of the world, overconfidence may prevent information aggregation with a strictly positive probability that does not vanish in large elections.

[23] Equilibrium analysis is very cumbersome if individuals are aware of others' overconfidence, but we prove a result in the case where the type space is finite and $\bar{q} = 1$ at the end of the proof of Proposition 2 in Online Appendix A.3.

Example 2: Let $\pi = 0.5$, and assume that there are three types of individuals in terms of competence: high type (H), medium type (M), and low type (L), where $\Pr(H) = \frac{1}{6}$, $\Pr(M) = \frac{1}{3}$, and $\Pr(L) = \frac{1}{2}$. Let $v \in \{l, h\}$ denote the state of media veracity. The signal precision for each type in each state of the world and each state of media is given in the table below:

	H	M	L
$\Pr(s_i = \alpha \mid S = A \cap v = h)$	1	0.9	0.8
$\Pr(s_i = \alpha \mid S = A \cap v = l)$	0.6	0.5	0.4
$\Pr(s_i = \beta \mid S = B \cap v = h)$	1	0.9	0.8
$\Pr(s_i = \beta \mid S = B \cap v = l)$	0.6	0.5	0.4

In this example, S and v are independent, and $\Pr(S = A) = \Pr(v = h) = 0.5$. In the unbiased electorate, information will aggregate; that is, the correct policy will be chosen with a probability that goes to one as N goes to infinity in both states. However, if $v = l$, and sufficiently many (for example, more than 2/3 of) low type individuals are overconfident and perceive themselves as high type, then in at least one state of the world the wrong policy will be chosen with a probability that goes to one because the expected relative turnout rate for the correct policy will certainly be lower than 50 percent due to overconfidence (see the proof in Online Appendix A.5). This holds regardless of whether or not individuals are aware of others' overconfidence, and despite the fact that every individual is on average more likely to be correctly informed than misinformed in both states of the world.

We now focus on the special case where the state of the world and the state of media veracity are perfectly correlated. In particular, one state is associated with high media veracity and the other with low media veracity; that is, $\Pr(S = A \cap v = l) = 0$ and $\Pr(S = B \cap v = h) = 0$ using the notation of Example 3. One justification for this scenario is that it may be more difficult to aggregate information in one state of the world due to reasons particular to that state. For example, Krishna and Morgan (2011) argues that such asymmetries arise naturally in circumstances where one candidate is the incumbent (of unknown type) because while most policy failures are often immediately evident, policy successes are only evident with the fullness of time, and thus, signals in the case where the incumbent is a good type are less precise (in Krishna and Morgan (2011) individuals are not overconfident, and signal precision q does not differ across individuals).

Another argument that relates more to the motivation behind misinformation dissemination is some form of media capture or public opinion manipulation by a political, corporate or other special interest using sponsored news, paid articles, and hidden advertisements. For example, propaganda is routinely used by politicians to sway opinions and behavior. As another example, a third party (e.g., a special interest, lobby, or industry) that strictly prefers one of the two policies regardless of the state of the world may try to influence public opinion and policy makers using scientists, experts, NGOs, and news media in order to facilitate the implementation of their preferred policy. Indeed, the tobacco industry, oil and energy lobby, and the military industrial complex have extensively used these channels to disseminate "self-serving news" (see Kartal and Tremewan (2018) and the references therein for detailed examples). Since such news are innocuous in one state of the world and false in the other, the end result is increased obfuscation and reduced media veracity in one state of the world (but not in the other). We provide a simple example to illustrate how overconfidence may harm information aggregation in this scenario.

Example 3: Assume that there are two types: high type (H) and low type (L). The signal precision for each type in each state is as follows.

	H	L
$\Pr(s_i = \alpha \mid S = A)$	1	p
$\Pr(s_i = \beta \mid S = B)$	1	$1 - p$

where $p \in \{1, 0.5\}$. First, assume that $p = 0.5$. In this case, information will always aggregate (unless all high types are underconfident and abstain). But what if a special interest group that strictly prefers policy a regardless of the state of the world can increase p from 0.5 to 1? We will now show that the special interest group can benefit from increasing p to 1 but *only* in sufficiently overconfident electorates. Note that a low type individual is essentially *uninformed* in either case whether $p = 1$ or $p = 0.5$. However, these are two distinct forms of "uninformedness," and their consequences may starkly differ in overconfident electorates unlike in unbiased electorates. If $p = 0.5$, then roughly equal numbers of low type individuals observe α and β signals in large elections, whereas if $p = 1$, then all low type individuals observe an α signal (for example because the special interest group inundates media outlets targeted to low type individuals with sponsored news and experts promoting policy a). In particular, if $p = 1$, and there are sufficiently many low type individuals sufficiently many of whom are overconfident, then policy a is chosen in both states with a probability that goes to one, as desired by the interest group. This holds regardless of whether individuals are aware or unaware of others' overconfidence.

This setup does not require the interest group to know the state of the world but simply to disseminate as many favorable news for its preferred policy as possible. As a result, the share of misinformation increases and the average signal precision decreases in only one state of the world (the average signal precision may even increase in the other state, which does not affect any of the results below).

In order to characterize equilibria in a more formal yet still tractable setting, we model signal precision of each individual as a function of media veracity v_S in state $S \in \{A, B\}$ and individual skill q, which is an i.i.d. draw from a common skill distribution $F[\underline{q}, \bar{q}]$ (by abuse of notation). The respective signal precision for individual i in state A and state B is given by $g(q_i, v_A) = q_i$ and $g(q_i, v_B) = m(q_i)$, where $m(q)$ is increasing in q. We assume that $m(q) \leq q$ for every q and $m(q) < q$ for some q representing the media influence of an interest group that strictly prefers policy b.[24] The function $m(q)$ is common knowledge. While individual i observes q_i (or rather perceives it as $p_i(q_i)$), i does not know with certainty whether the true precision is q_i or $m(q_i)$ as i does not know the state of the world. Equilibrium characterization in this setting closely resembles that in Lemma 1. Let q^a and q^b be as defined in (2) and (3) respectively. For brevity in notation we assume that $p_i(q_i) = q_i$. Equilibrium strategies are as follows.

- Individual i votes for a if and only if either i's signal is α and $\frac{q_i}{q_i+1-m(q_i)} \geq q^a$ or i's signal is β and $\frac{m(q_i)}{m(q_i)+1-q_i} \leq 1 - q^a$.
- Individual i votes for b if and only if either i's signal is β and $\frac{m(q_i)}{m(q_i)+1-q_i} \geq q^b$ or i's signal is α and $\frac{q_i}{q_i+1-m(q_i)} \leq 1 - q^b$.

As $\frac{q}{q+1-m(q)}$ and $\frac{m(q)}{m(q)+1-q}$ are monotone increasing in q, equilibrium voting behavior is once again characterized by cutoffs.[25] Consider as an example the case where $m(q) = q^x$ and $x > 1$. As x increases, average signal precision drastically falls in state B, and eventually $\int_{\underline{q}}^{\bar{q}} m(q)dF < 0.5$ will hold. Note that $\int_{\underline{q}}^{\bar{q}} m(q)dF < 0.5$ can hold even if $\pi\underline{q}+(1-\pi)m(\underline{q}) \geq 0.5$ (that is, on average across the two states every individual is more likely to be correctly informed than misinformed). This represents the generalization of the simple setting in Example 3 above. If $\int_{\underline{q}}^{\bar{q}} m(q)dF < 0.5$, then a majority will observe an α signal in state B, and thus, making the correct group decision in state B requires many people with low q to

[24] The preference of the interest group for policy b is without loss of generality. If the interest group prefers policy a, then we set $g(q_i, v_A) = m(q_i)$ and $g(q_i, v_B) = q_i$.

[25] The prior π is accounted for in q^a and q^b as before.

refrain from following their α signal. However, if too many people are overconfident, then too many of those will simply vote according to their signals, and policy a will be chosen in both states with a probability that goes to one, as desired by the "source" of the low media veracity in state B. As mentioned above, this setup does not require the interest group to know the state of the world but simply to sponsor as many self-serving news as possible. One caveat is that for the case where $\bar{q} = 1$ in Proposition 3, we assume that every i is an a-partisan with probability $\frac{p}{2}$ and a b-partisan with probability $\frac{p}{2}$ where $p > 0$ (but possibly small). We need this assumption for tractability.

Proposition 3 "A third party (e.g., a special interest, lobby, industry, or political party) may benefit from low media veracity and high voter overconfidence." *Assume that π is bounded away from 0 and 1, and that $m(\bar{q}) > 1 - m(\underline{q})$. If media veracity in state B is low enough so that $\int_{\underline{q}}^{\bar{q}} m(q)dF < 0.5 < \int_{\underline{q}}^{\bar{q}} q dF$ holds, and the electorate is sufficiently overconfident, then the probability that policy a is chosen goes to one in both states as N grows without bound, whereas in an unbiased electorate the correct policy is chosen with a probability that goes to one in both states.*[26]

(iii) This model relaxes the assumption that individuals observe statistically independent signals. Instead, there are finitely many news sources generating statistically independent news signals, and each individual obtains one signal from one news source. As a result, in large elections numerous individuals will receive an identical news signal from the same news source. News sources differ in their quality. More specifically, there are $n > 0$ reliable high quality news sources and $m > 0$ unreliable low quality news sources. A reliable news sources provides one independent high quality signal $s_h \in \{\alpha, \beta\}$ with a precision of $q_H > 0.5$ such that $\Pr(\alpha|A) = \Pr(\beta|B) = q_H$. An unreliable source provides one independent low quality signal $s_l \in \{\alpha, \beta\}$ with a precision of $q_L \leq 0.5$ such that $\Pr(\alpha|A) = \Pr(\beta|B) = q_L$. The term q_i $(1 - q_i)$ denotes the probability with which individual i's signal s_i comes from one of the high-quality (low-quality) news outlets, with each high-quality (low-quality) outlet being equally likely to provide s_i in that case. Hence, as before, the higher the magnitude of q_i, the higher the probability of observing correct news. As before, q_i is an independent draw from a common distribution $F[\underline{q}, \bar{q}]$. Possible interpretations of this variation are various forms of media censure or capture as well as increased media ownership concentration as these are

[26] Equilibrium analysis is once again very cumbersome if individuals are aware of others' overconfidence, but we prove a result in the case where the type space is finite and $\bar{q} = 1$ at the end of the proof of Proposition 3.

likely to be associated with a decrease in the number of news sources generating statistically independent reports.[27] We now provide a simple example to illustrate how overconfidence may harm information aggregation in this scenario.

Example 4: Assume that $n = 5$ and $m = 5$, with $q_H = 0.8$ and $q_L = 0.5$. There are two types of individuals in terms of competence: low type individuals with $q_i = 0$ and high type individuals with $q_i = 1$ where $F(0) = 0.7$. The equilibrium in which only high type individuals vote and the rest abstain is optimal in the limit. If for example 50 percent of low type individuals believe that they are high type, then the rate of information aggregation in large elections is about 16 percentage points lower relative to that in an unbiased electorate because overconfident low type voters reduce the rate of information aggregation (see the proof in Online Appendix A.5).

The Bayesian Nash equilibrium consists of cutoffs also in this model, but the characterization of equilibrium behavior is much more involved than in the previous variations due to the correlated nature of signals and therefore relegated to the proof of Proposition 4 in Online Appendix A.3. For tractability, we assume that $q_L = 0.5$, and that n and m are odd numbers.

Note that even when $\underline{q} < 0.5$, every individual is more likely to be correctly informed than misinformed in either state in this model because $q_i q_H + (1 - q_i) q_L > 0.5$ for every $q_i > \underline{q}$. In the independent-signal model, this is a sufficient condition to ensure that the correct policy is chosen in both states in large elections regardless of the extent of overconfidence. However, correlation among signals results in correlated and thus possibly large mistakes due to overconfidence. Therefore, the negative impact of overconfidence on information aggregation may not vanish in the limit even though every individual is more likely to be correctly informed than misinformed in both states. While we can characterize the optimal equilibrium of this model with a continuous type space, in order to prove the inefficiency result we need to switch to a discrete type space as the analysis is too cumbersome otherwise.

Proposition 4 (Impact of media concentration/censure with high voter overconfidence) *Assume that $q_L = 0.5$ and that $\pi \in (1 - q_H, q_H)$ but bounded away from q_H and $1 - q_H$.* **(i)** *If $q_H = 1$ and $\mathbb{E}(q) < 0.5$, an unbiased electorate makes the correct decision with a probability that goes to one in both states as N grows without bound, whereas in a*

[27] Bagdikian (2004) wrote that the number of corporations controlling most of the media in the US decreased to 5 from around 50 in 1983 and noted that "this gives each of the five corporations and their leaders more communications power than was exercised by any despot or dictatorship in history."

sufficiently overconfident electorate the limiting probability is strictly lower than one. **(ii)** *If $q_H < 1$, an unbiased electorate makes the correct decision with a probability that goes to $\sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q_H^i (1-q_H)^{n-i}$ in both states as N grows without bound, whereas in a sufficiently overconfident electorate the limiting probability is strictly lower than that if $\bar{q} > \frac{n}{n+1}$ and $\mathbb{E}(q) < \frac{n}{n+1}$ hold.*

To reiterate, even if $\underline{q} < 0.5$ and $\mathbb{E}(q) < 0.5$ every individual is more likely to be correctly informed than misinformed in either state since $q_L = 0.5$. However, unlike in a model with independent signals, this is not enough to suppress the effect of overconfidence in large elections. We can also show that all else equal the negative effect of overconfidence is maximized at $m = 1$. Thus, even the concentration of media ownership among low quality media outlets (i.e. low m) can be highly problematic in the presence of overconfidence. In unbiased electorates, however, m has no impact on information aggregation in arbitrarily large elections as implied by Proposition 4.[28]

Finally, note that it is possible to combine the last two models with correlated news reports and a third party that disseminates self-serving news. The third party may either own a number of (low-quality) outlets or have a close relationship to them. Thus, especially q_L may depend on the state of the world (assuming high quality outlets are more difficult to influence). For example, if the third party strictly prefers policy a, then news reports promoting a could make q_L high in state A and low in state B making overconfident individuals' votes highly tilted towards policy a as in Example 3. As a result, the joint impact of media concentration, third-party interference, and voter overconfidence can be highly harmful. For instance, the inefficiency caused by overconfidence in Example 4 increases if we assume that low quality outlets are such that $q_L = 0.8$ in state A and $q_L = 0.2$ in state B due to the interference of a special interest that strictly prefers policy a. Then, all else equal the negative effect of overconfidence is 28 percentage points instead of 16 with $q_L = 0.5$ in both states (in particular, decision making accuracy falls by 59.9 percentage points in state B while increasing by 3.8 in state A relative to the unbiased case).

[28]The condition in part (ii) of Proposition 4 may seem to suggest that overconfidence is more likely to be a problem as n (i.e., the number of reliable news sources) increases, but there are two countervailing forces. While it is true that the sufficient condition becomes easier to satisfy as n increases, the information aggregation benefit of increased n will likely mitigate the impact of overconfidence.

3 Experiment

The objective of our experiment is to test some of the implications of our theory in environments with high levels of misinformation and overconfidence. Our study and results also relate to the experimental literature on individual overconfidence by showing that collective decision making with overconfidence can result in outcomes that are much more extreme than the aggregate of individual decisions with overconfidence. While collective decision making can mitigate or even fully correct for the effect of overconfidence in many settings, it may also generate strongly inefficient outcomes in settings with high levels of false information and overconfidence as our theory indicates.

3.1 Experimental Design and Predictions

The experiment consists of several parts and two within-subject treatments. We first explain the details of our main treatment, which is largely based on the voting game analyzed in Section 2.1 with $\underline{q} < 0.5$ and perception biases. Subjects make choices in groups of 24 (that is, $N = 24$). The state of the world is either *Red* or *Blue* (described as "group color" in the experimental instructions), and either state is ex-ante equally likely. After privately observing a signal, which is red or blue, subjects vote for either red or blue, or abstain. The group decision is determined by majority voting, and group members receive a monetary reward if the group decision matches the group's color (i.e. the state of the world).

We implement individual signal accuracy q_i, correct news, fake news, and the subjective perception of accuracy $p_i(q_i)$ as follows. In the experiment, subjects take a quiz on math and logic puzzles before learning about the voting phase. The quiz is incentivized as each correct answer in the quiz is rewarded. Subjects are not informed about their performance in the quiz until the end of the experiment. During the voting phase, subjects are explained that the signal they receive regarding the state depends on their quiz score; that is, a group member observes a signal that matches the true state (correct news) if and only if the member's quiz score places him/her in the top 1/3 of all the subjects in the same experimental session. So, signal accuracy q and the subjective perception of accuracy $p(q)$ depend on the (perceived) ability to perform well in the quiz. To be more precise, q equals the ex-ante probability of being in the top 1/3 of all the subjects in the same experimental session. Since subjects do not learn whether or not they are in the top 1/3 until the end of the experiment, they must form a belief about it, which is precisely $p(q)$. The quiz is followed by an incentivized belief elicitation task regarding the likelihood of placing in the top 1/3 of all the subjects in the same session.[29]

In the design, we chose to use relative performance to determine the signal of a subject due to following reasons. First, it enables us to use the techniques developed in Benoit and Dubra (2011) and Benoit, Dubra, and Moore (2015) in order to test for *true* overconfidence. Second, it enables us to fully control the fraction of correct news in a large laboratory election setting. If q was based on absolute performance, it would be more difficult to finetune the quiz in a way that many of our sessions have a sufficiently large number of subjects who observe an incorrect signal. This is important for our design because we would like to implement $\underline{q} < 0.5$. Third, the theoretical mechanism of the swing voter model concerns relative competence, not absolute competence: for example, given $q_i = 0.6$, i should definitely vote if $q_j = 0.5$ for every $j \neq i$, whereas i should likely abstain if $q_j = 1$ for $j \neq i$.[30]

In the experimental design, we impose *informative voting*; that is, we do not allow a subject to vote for red if the subject's signal is blue, and vice versa. Our theoretical predictions take this restriction into account. We impose informative voting in order to reduce noise in the data. As will become clearer in the next section, we have an elaborate experimental design with multiple phases and multiple belief elicitation tasks. In order to be able to attribute an inefficiency in decision making to the correct source (overconfidence versus not understanding the incentives in the game), we wanted to make the signal structure and the voting choice as simple as possible so that subjects have a very clear understanding of the voting game and belief elicitation tasks.

Under the assumption of informative voting, a Bayesian Nash equilibrium consists of a cutoff belief regarding one's success in the quiz: those who are sufficiently confident that their quiz score places them in the top 1/3 should vote for the color of their signal, and the rest should abstain and delegate the voting decision. With subjective beliefs, we cannot provide point predictions for the equilibrium cutoff and the probability that the group makes the correct decision as we do not know the true distribution of q. However, the cutoff belief has to be higher than 50 percent in the optimal equilibrium. This forms the basis of the first part of Hypothesis 1. More generally, the best response of subject i for some fixed strategy of others is always characterized by a cutoff belief such that i votes for the color of i's signal if i's belief is above the cutoff and abstains otherwise (recall that voting for the color that

[29] Subjects learn about the belief elicitation task only after completing the quiz.

[30] Also, we chose 1/3 as the fraction of correct news, because once the fraction of correct news exceeds 1/2, then strategic abstention is not necessary to make the correct group decision, and overconfidence should not have a harmful effect on information aggregation.

does not match one's signal is not allowed). We can compute a best-response cutoff for each subject as we elicit subjects' beliefs regarding the probability that a randomly selected group member chooses to vote and the probability that a randomly selected "voter" is in the top 1/3 of all the subjects in the same session.[31] (It was made clear to subjects that the randomly selected group member or voter was always someone else in the room.) These elicited beliefs are enough to estimate a best-response cutoff for each subject. This forms the basis of the second part of Hypothesis 1.

Hypothesis 1 (i) Weak test for voter rationality: *Subjects who choose to vote state a belief greater than 50% regarding the likelihood of being in the top 1/3.* **(ii) Strong test for voter rationality:** *Subjects who choose to vote state a belief which is greater than their respective (computed) cutoff beliefs.*

While Hypothesis 1 relates to rationality of observed turnout (however allowing for biased beliefs), Hypothesis 2 concerns the efficiency of turnout when subjects' beliefs are biased, assuming that subjects are otherwise rational. As discussed in our theoretical model, biased beliefs result in a deviation from the best response behavior, and this deviation results in an inefficiency (relative to the optimal unbiased equilibrium outcome). As mentioned above, we cannot provide a point prediction for the probability that the group makes the correct decision as we do not know the true distribution of q. However, the probability that the group decision is correct must exceed 50 percent in every equilibrium if the group consists of rational and unbiased agents.

Hypothesis 2 (Inefficiency of Overconfidence) *If the electorate is highly overconfident (but rational otherwise), the probability of a correct group decision is lower than* 50 *percent.*

Hypothesis 2 is a very stringent test regarding the negative effect of biased beliefs because 50 percent is a distribution-free lower bound for the probability of a correct group decision with unbiased and rational subjects.[32] As an example, if individuals who are in the top 10% of the *belief* distribution in an unbiased population (regarding the probability of being in the top 1/3) has an expected probability $y > 0.5$ for being in the top 1/3,

[31] For the cutoff computation, we assume that strategies of other members are *symmetric* with respect to signals, which implies that the best response of every subject is also symmetric. This assumption is arguably not very restrictive given that the environment is symmetric. Moreover, relaxing the assumption would require us to elicit beliefs based on each possible state separately, which would be too cumbersome.

[32] On a side note, if we removed the informative voting restriction in theory and in the experimental design, the data we obtained would be compared to an accuracy lower bound of 94.5%, because in that case it is possible to show that the theoretical accuracy of the optimal equilibrium must be higher than 94.5%.

then the theoretical probability of a correct group decision will be very close to or exceed y in the optimal equilibrium (e.g., if $y = 0.6, 0.7$, and 0.8, then the probability of a correct group decision in the optimal equilibrium can be shown to be greater than 0.6, 0.7, and 0.8 respectively).

Importantly, Hypothesis 2 is robust to assuming that instead of the equilibrium cutoff subjects use a heuristic cutoff such as 50 percent (e.g., a subject votes if the subject's confidence level for being in the top 1/3 is greater than 50 percent and abstains otherwise). The probability that the group decision is correct must also exceed 50 percent under this cutoff heuristic if the group consists of unbiased individuals. Therefore, Hypothesis 2 is immune to the findings in Esponda and Vespa (2015) that subjects have difficulties with understanding the logic of pivotality and therefore make suboptimal voting decisions.

3.2 Experimental Protocol

The experiment was conducted at the experimental laboratory of the Vienna Center for Experimental Economics (VCEE) at the University of Vienna. Subjects were recruited from the general undergraduate population. The experiment consists of several parts. Every session began with instructions on a belief elicitation task using quadratic scoring rule, followed by incentivized belief elicitation exercises (unrelated to the voting game) because we use belief elicitation tasks in various parts of the experiment, and we wanted the participants to be familiar with the task in order to reduce noise in the data.

After this part, participants take the quiz on math and logic puzzles. The quiz consists of 20 questions, subjects are informed that each correct answer is rewarded (by 30 Eurocents), and they have 10 minutes to work on the quiz. We asked simple questions as prior experimental research suggests that easy tasks are conducive to overconfidence (i.e., overplacement), and thus the negative effect of overconfidence on information aggregation is more likely to be borne out with an easy quiz.[33] The average number of correct answers in the quiz is 16.8 and the 33rd percentile is at 18. One relevant question is the following: are policy issues simple for the average voter? Arguably, many of them are not — however, it suffices if the average voter "perceives" them to be so, and there is evidence for this from the literature showing that people have strong policy opinions while being somewhat

[33] Prior work suggests that populations tend to exhibit overconfidence on easy tasks and underconfidence on difficult tasks (see, e.g., Moore and Cain, 2007; Moore and Healy, 2008; and Benoit, Dubra, and Moore, 2015).

informed, uninformed or misinformed (see, for example, Kuklinski et al. (2000) and Flynn et al. (2017))

After the quiz is over, the first incentivized belief elicitation task related to our main treatment is given: we ask subjects to report a percentage (between 0 and 100) regarding the likelihood that their quiz score places them in the top 1/3 of all the subjects in the same session. This is followed by the voting phase. The voting phase starts with an initial pseudo-control/teaching treatment, which consists of 15 rounds. The main aim of this treatment is to teach the voting game using an *objective* information structure and analyze whether participants understand that they should abstain when their signal precision is low. In each round, each participant observes a private signal that is either red or blue as in the main treatment, but participants are informed about the "true" precision of their signal for that round. Put differently, signal precision q is an objective value in this treatment. Participants also know that the signal precision of every other group member is an iid uniform draw. After every round, participants receive detailed feedback about the round's outcome including the accuracy of the group decision and whether or not their signal was correct. We refer to this treatment as the OBJ treatment. After this part, the main treatment starts. In this treatment, participants learn that whether or not their signal matches the true state depends on their performance in the quiz. Thus, the signal precision of each participant has a subjective value, which we elicited using an incentivized belief elicitation task after the quiz as mentioned above. We refer to this treatment as the SUBJ treatment.

SUBJ treatment involves two belief elicitation tasks related to other group members as mentioned above. The elicited beliefs are about the probability that a randomly selected group member chooses to vote and the probability that a randomly selected voter is in the top 1/3 of all the subjects in the same session; in both cases, it was made clear to subjects that the randomly selected member (or voter) was someone else in the room. SUBJ treatment consists of 6 rounds, and these 6 rounds were divided into two parts. In the first part consisting of 5 rounds, participants were only asked about the probability that a randomly selected group member chose to vote after making their voting decision.[34] From those rounds, one round was randomly selected for payment. In the second part which consisted of only one round, we elicited both beliefs regarding the probability that a randomly selected group member chooses to vote and the probability that a randomly selected voter is in the top 1/3 of all the participants in the same session. This round was fully paid including the belief

[34] This belief elicitation task was also given in the OBJ treatment.

elicitation tasks—so it was a high-stake final round. Subjects did not receive any feedback after making their choices in the SUBJ treatment (except at the end of the experiment).

We had a total of 6 sessions (i.e., 6 laboratory electorates) with 144 subjects. As mentioned above, subjects did not receive any feedback after making their choices in the SUBJ treatment. In a follow-up treatment, which also had 6 sessions and 144 subjects, subjects received aggregated feedback after each round regarding the accuracy of the group decision (but not regarding the true state as this would fully reveal their placement in the quiz) and the total number of group members who voted. We refer to this treatment as the SUBJ+ treatment. These sessions with the SUBJ+ treatment had exactly the same protocol as before (e.g., they included the OBJ treatment) with the exception that there was aggregate feedback after every round in the SUBJ+ treatment.

A total of 288 undergraduates participated in 12 experimental sessions. All sessions were conducted through computer terminals, using a program written in Z-Tree (Fischbacher, 2007). In each phase of a session, the experiment only started when all subjects had correctly answered a set of control questions that tested whether they understood the instructions. The average payoff per subject was about €18, and each session lasted between 100 minutes and 2 hours. The instructions for the treatments can be found in the Online Appendix.

4 Experimental Results

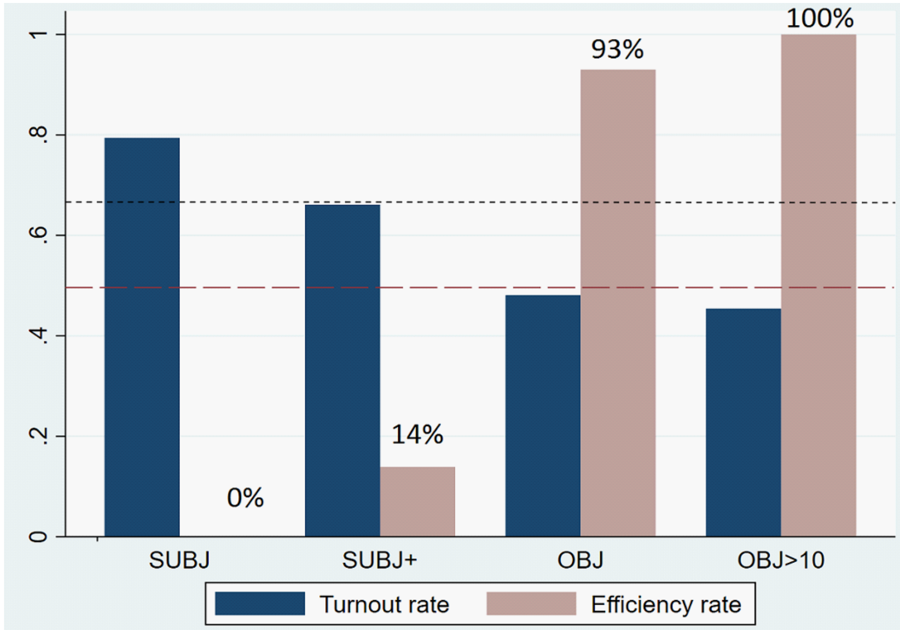
The presentation of experimental results are organized as follows. We start with a discussion of aggregate results on turnout and efficiency (i.e., the fraction of correct group decisions). Statistical tests are based on nonparametric statistics, and these tests are two-sided and conducted at the session (electorate) level. In addition, we use random effects probit estimations to analyze turnout behavior at the individual level (see Online Appendix B). Finally, we discuss the results of the follow-up treatments with feedback in Section 4.3.

4.1 Efficiency of Group Decisions, Turnout, and Elicited Beliefs

Figure 1 shows the percentage of correct group decisions (referred to as efficiency in the figure) and the average turnout rate across treatments. Out of 36 group decisions made in 6 sessions over 6 rounds in the SUBJ treatment, not a single group decision was correct as indicated in Figure 1.

Result 1: *The percentage of correct group decisions in the SUBJ treatment is 0%, which is*

Figure 1: Average Turnout and Efficiency by Treatment



Notes: The long dashed red line at 0.5 represents the theoretical efficiency benchmark against which the percentage of correct group decisions is compared in SUBJ and SUBJ+ and the short dashed blue line at 0.667 represents the rate above which turnout rate is *excessive* and certainly results in *incorrect* group decisions. OBJ>10 refers to OBJ data after dropping first 10 rounds.

significantly lower than the 50% threshold according to both Wilcoxon signed-rank test and sign test with respective p-values of 0.0277 *and* 0.0313.[35]

We explore the reason for this drastic failure of information aggregation analyzing turnout behavior and elicited beliefs of subjects. We first consider aggregate turnout. Figure 1 shows that the average turnout rate is slightly lower than 80% in the SUBJ treatment. Note that by design a turnout rate that exceeds 66.7% immediately implies that the group decision is wrong because only one third of group members observes correct news. Thus, group decisions are wrong because voter turnout in SUBJ is consistently very high in all six sessions with an average turnout rate of 80.6% in the final round (which is a high-stake round) and an average turnout rate of 79.4% across all rounds. Moreover, in every session, the average turnout rate across rounds is 75% or higher. Figure 2 plots the turnout rate over time in every session.

[35]Since we have six (independent) pairs of observations, the two-sided Wilcoxon signed-rank test and the two-sided sign test can never reject a hypothesis with p-values lower than 0.0277 and 0.0313, respectively.

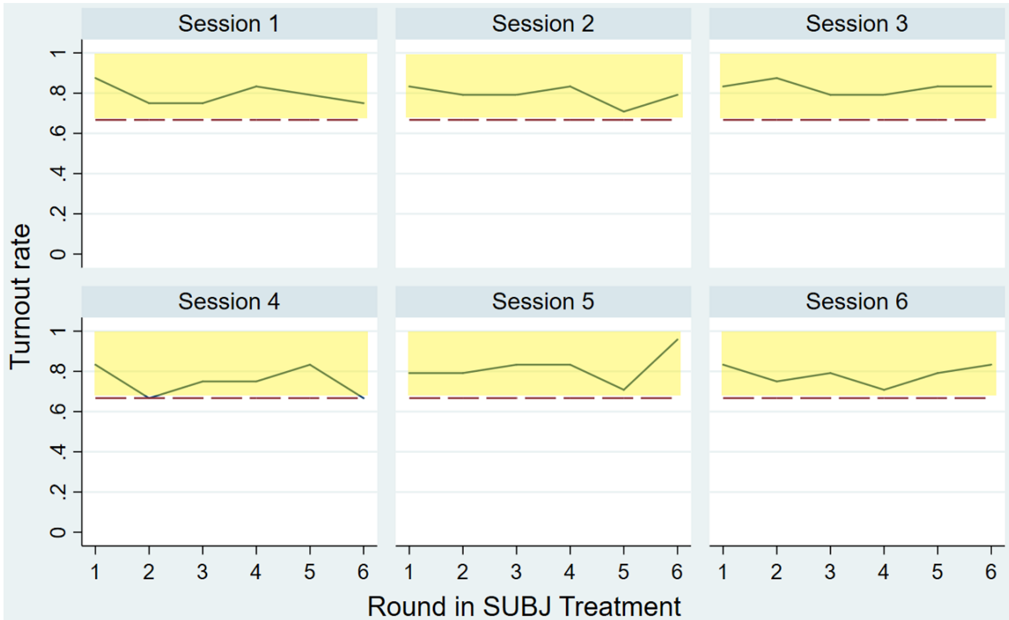
A substantial fraction of turnout can be rationalized analyzing the elicited beliefs. Applying the weaker test in Hypothesis 1, we find that across all rounds less than 6.3% of voters have stated beliefs lower than 50% regarding their chances of placing in the top 1/3 of all the subjects in the same session. As discussed above and stated in Hypothesis 1, we can also apply a highly stringent rationality criterion deriving the best response of every subject using the subject's elicited beliefs regarding (i) the probability that a randomly selected group member chooses to vote, and (ii) the probability that a randomly selected voter is in the top 1/3. We then compare each voter's behavior to the theoretical best response of the voter given the computed best-response cutoff belief and the belief we elicit from each subject regarding their likelihood of placing in the top 1/3 of their session. In the final round, only 10 percent of voters behave in a way that is inconsistent with the best response behavior we predict for them.[36] These suggest that turnout behavior in the SUBJ treatment is to a considerable extent consistent with rational behavior as predicted in Hypothesis 1. We provide additional support for rationality of voter behavior using random effects probit estimations in Online Appendix B.

Result 2: *Voter behavior in the SUBJ treatment is to a large extent consistent with rationality on the basis of elicited beliefs.*

Figure 1 shows that the turnout rate in the OBJ treatment is moderate: it is only 48.1% and much lower than that in SUBJ although the expected fraction of correct news is also higher in OBJ than in SUBJ (50% in OBJ versus 33.3% in SUBJ). There is also learning towards abstention in the OBJ treatment as turnout rate falls in every session if, for example, we compare the turnout rate in the first five rounds to that in the last five rounds. In particular, the average turnout in the first five rounds is 51.3%, and this falls to 45.4% in the last five rounds (the decline in turnout is statistically significant according to both Wilcoxon signed-rank test and sign test at $p < 0.05$). Also, voting is very rare

[36] For 25 voters (out of a total of 116 voters in the final round), it was not possible to estimate a cutoff as each of these voters indicated that a randomly selected group member would choose to vote with 100 percent probability. A voter cannot be pivotal in such a case by experimental design (because only 8 individuals observe a correct signal, and it is not possible to vote against one's signal), and thus, it is mathematically not possible to find a cutoff. However, of those 25 voters, 19 indicated that they are 100 percent confident that they are in the top 1/3, and thus it is a dominant strategy to vote for them regardless of a cutoff. Therefore, we subtract only 6 subjects from a total of 116 voters due to not being able to compute a cutoff. Since 11 voters indicated beliefs regarding their quiz performance lower than the cutoff belief which we estimated for them, we arrive at 10 percent=11/110.

Figure 2: Turnout in SUBJ Treatment by Session



Note: The solid blue line represents the turnout rate and the shaded area above the horizontal dashed line at 0.667 represents *excessive* turnout rates which result in *incorrect* group decisions with certainty.

in OBJ (similar to that in SUBJ) once the signal precision falls below 50%. Finally, the percentage of correct group decisions is on average 93.33% across all rounds and increases to 100% in the last 5 rounds as shown in Figure 1. While we cannot directly compare SUBJ and OBJ treatments as the distributions of signal precisions are not identical (in particular, the expected fraction of correct news differ), OBJ treatment is still a useful benchmark to show that if subjects have objective information about the quality of their signals, then they delegate the voting decision frequently (even more so when they are experienced), average turnout is far from excessive, and group decision making is efficient.

Result 3: *Voter behavior in the OBJ treatment is to a large extent consistent with rationality.*

4.2 The Effect of Overconfidence on Group Decisions

To document the effect of subjective beliefs on turnout and efficiency, we use the following test based on the work by Benoit and Dubra (2011) and Benoit, Dubra and Moore (2015). We disregard abstainers and compare at the session level the percentage of voters who end up being in the top 1/3 to the average elicited belief among voters regarding their chances of being in the top 1/3 of all the subjects in the same session. If there is a systematic

Table 1: Voters' elicited beliefs about placement in the top 1/3 vs voters' actual placement

Session	**Average probability voters assign to own placement in top 1/3**	**Percentage of voters placed in top 1/3**
1	83.9%	33.3%
2	85.05%	36.8%
3	88.6%	40%
4	92.5%	37.5%
5	82.7%	34.8%
6	80.8%	30%

(statistically significant) upward bias in elicited beliefs relative to the actual placements, then there is evidence for overconfidence among voters, which explains the high turnout.

For this part, we focus on the last round of the SUBJ treatment, which was a high stake round—however, the conclusion does not change if we analyze separately every other round in that treatment. Table 1 shows the average probability voters assigned to being in the top 1/3 of their session and the actual fraction of voters who are in the top 1/3 in every session. In every session, there is a substantial discrepancy between elicited beliefs and actual placements, and the difference is significant according to both Wilcoxon signed-rank test and sign test at 5% level with respective p-values of 0.0277 and 0.0313.[37] On average the elicited belief among voters regarding the likelihood of being in top 1/3 is 85% but the percentage of voters eventually placed in top 1/3 is only 35%. As previous section has shown that voter behavior can be rationalized to a substantial extent on the basis of elicited beliefs, we conclude that overconfidence is responsible for a large share of the excessive turnout we observe, which in turn prevents groups from making the correct decision (recall that only one third of the session observes correct news, and thus, a turnout rate that exceeds 66.7% surely results in a wrong decision). Finally, note that in every session, those who are placed in top 1/3 state higher beliefs (93%) than those who are not (75.8%), which is significant at 5% level according to both Wilcoxon signed-rank test and sign test. Thus, despite prevalent overconfidence, higher placement is still associated with higher beliefs. Put differently, those who state higher beliefs are more likely to observe the correct signal.

Combining all the findings discussed above regarding turnout, elicited beliefs and the

[37] Some experimental studies have found that men are more (likely to be) overconfident than women. We do not observe this in our experiment.

accuracy of group decisions, we conclude that we find support for Hypothesis 2.[38]

Result 4: *Subjects who choose to vote are statistically and economically significantly overconfident, which inflates turnout. In turn, excessive turnout generates 0% accuracy in group decisions.*

4.3 Follow-up Treatment: The Effect of Aggregate Feedback on Turnout and Group Decisions

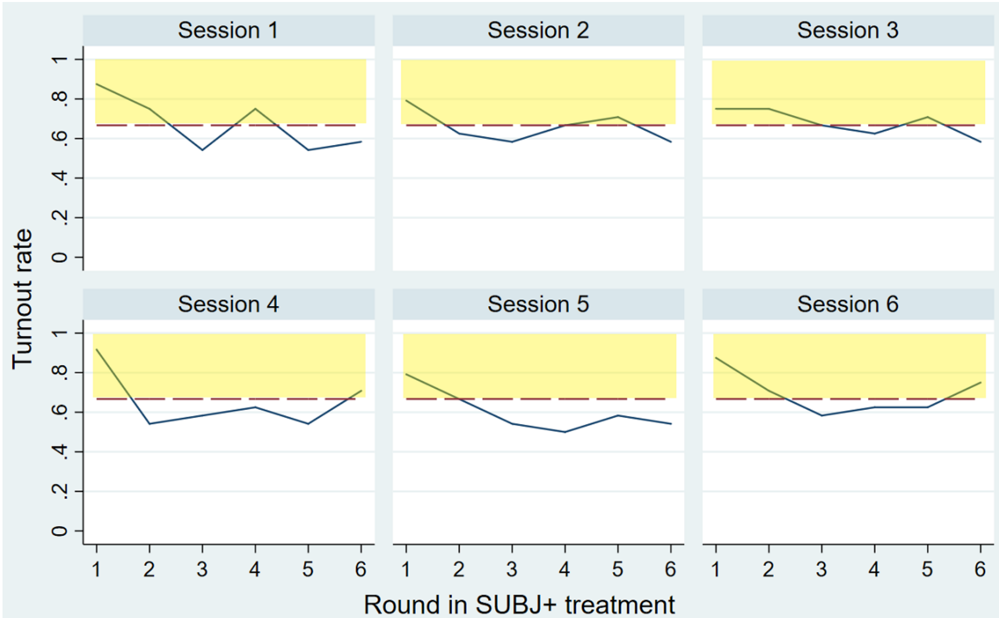
Out of 36 group decisions made in 6 sessions in the SUBJ+ treatment, only 5 decisions were correct. Even though efficiency increased from 0% to 13.9% with feedback, we again find support for Hypothesis 2 according to a Wilcoxon signed-rank test at 5% level ($p > 0.1$ with sign test).

It turns out that aggregate feedback resulted in a sizeable reduction in turnout in comparison to SUBJ: the average turnout rate across all rounds is 66.1%. (see Figure 1). Turnout also fell across rounds (see Figure 3 for the turnout rate across rounds in all sessions). For example, comparing the first and the final rounds with subjective signals, turnout rate fell by a quarter from 83.33% in the first round to 62.5% in the sixth round (this is significant according to both Wilcoxon signed-rank test and sign test at 5% level).

However, reduction in turnout failed to significantly improve efficiency because the increase in the share of voters in top 1/3 due to aggregate feedback was very limited (e.g., it increased from 36.66% in the first round to 41.11% in the final round). This is because some of the voters in top 1/3 also reduced their turnout rate with feedback, and overconfidence among those who vote proved once again rampant. More specifically, comparing the first and the final rounds, turnout of voters in top 1/3 and turnout of voters placed below top 1/3 fell by 15.9% and 30.3% respectively. Moreover, as in the previous treatment without feedback, there is a consistently high discrepancy between the average elicited belief among voters regarding the likelihood of being in top 1/3 and the percentage of voters eventually placed in top 1/3 (85.3% versus 39.8%), which does not change over time and is significant at 5% level. Thus, despite aggregate feedback resulting in a sizeable decline in turnout, group decision making failed to significantly improve due to aggregate feedback making some voters in the top 1/3 less confident, and due to rampant overconfidence among voters.

[38] See also Online Appendix B for probit regressions, which provide a more detailed analysis of turnout behavior and additional support for our claim that individual voter behavior is to a large extent consistent with rationality.

Figure 3: Turnout in SUBJ+ Treatment by Session



Note: The solid blue line represents the turnout rate and the shaded area above the horizontal dashed line at 0.667 represents *excessive* turnout rates which result in *incorrect* group decisions with certainty.

5 Related Literature

There is an extensive literature analyzing the impact of overconfidence on individual decision-making as well as markets.[39] Different from most of the previous literature, we focus on the effect of overconfidence on voting and collective decision making.[40] In our setup, individuals differ in their competence and may choose to abstain. Therefore, our model is foremost related to Feddersen and Pesendorfer (1996). There have been various theoretical extensions of Feddersen and Pesendorfer (1996), such as Feddersen and Pesendorfer (1999), McMurray (2011), Herrera, Llorente-Saguer, and McMurray (2019a) as well as several experimental studies, which analyze among other things whether less-informed/uninformed individuals strategically abstain (e.g., Battaglini, Morton, and Palfrey (2010), Morton and Tyran (2011), Bhattacharya, Duffy, and Kim (2014), Elbittar et al. (2016), and Herrera, Llorente-Saguer,

[39] See among others Camerer and Lovallo (1999), Malmendier and Tate (2005), Barber and Odean (2000), and Benoit, Dubra, and Moore (2015).

[40] Our results also apply to decision making in committees. Committees typically consist of professionals, who may be overconfident regarding their expertise. According to Daniel Kahneman, "[O]verconfident professionals sincerely believe they have expertise, act as experts and look like experts. You will have to struggle to remind yourself that they may be in the grip of an illusion." (*New York Times*, 23 Oct, 2011)

and McMurray (2019b)).[41] Our study differs from this literature (with the exception of Elbittar et al. (2016)) in that people hold subjective (rather than objectively correct) beliefs about the quality of the news they receive. Elbittar et al. (2016) studies information aggregation under unanimity and majority rules in a common interest setting where each group member is initially uninformed about which of two alternatives is optimal but may choose to acquire a noisy information signal at a cost. Uninformed subjects only learn that each alternative is equally likely to be optimal. In contrast to the equilibrium predictions, many subjects vote rather than abstain despite being uninformed. Thus, differing from the findings of other experimental studies mentioned above as well as the OBJ treatment of our experiment, many subjects vote when it is optimal to abstain. The difference in the experimental findings of Elbittar et al. (2016) and the current study is likely because information acquisition is absent in our simple setup, which together with very detailed feedback we provide after every round in OBJ treatment may have made it much easier and quicker for subjects to understand the value of abstention. Elbittar et al. (2016) show that a biased voting model allowing for noisy priors (i.e., priors that randomly deviate from the objectively correct equal prior) can account for the observed prevalence of uninformed voting.

Morton, Piovesan, and Tyran (hereafter MPT, 2019) experimentally study information aggregation in a common interest setting without abstentions where voters may have misleading information. Subjects vote on the answers of various quiz questions some of which are known to the experimenter to be "misleading" (less than 40 percent of subjects got them right in a pretest). Results show that voters with higher cognitive ability (as measured by a cognitive reflection test) are more likely to vote for the correct answer. In addition, subjects are found to be overconfident in their ability to answer misleading questions correctly. Hence, MPT's experiment is related to ours in that subjects in both studies receive misleading information as a function of their cognitive ability, and voters in both cases tend to be overconfident in the sense that they do not recognize their information is inaccurate. However, there are important differences between the two studies. Concerning the experimental design, abstentions were not allowed in MPT, and the relation between voter information, cognitive ability, and overconfidence is tighter in the current study. Importantly, we provide a much more extensive theoretical analysis of information aggregation with overconfidence

[41] Our paper connects to a larger literature that studies information aggregation in two-alternative decisions (see, e.g., Austen-Smith and Banks, 1996; Duggan and Martinelli, 2001; Martinelli, 2006; Visser and Swank, 2007; Goeree and Yariv, 2011; Bhattacharya 2013; Bouton, Llorente-Saguer, and Malherbe, 2018).

than MPT.

Papers by Levy and Razin (2015) and Ortoleva and Snowberg (2015) investigating political behavior with correlation neglect are also related. Individuals' utilities from different policies depend on their (heterogeneous) preference parameters and an unknown state of the world, which is the same for all individuals. Each individual receives repeated information signals regarding the state of the world. Unlike in our model, individuals do not differ in their competence per se, but in the degree to which they underestimate the correlation between their signals.[42] Ortoleva and Snowberg (2015) show, theoretically and empirically, that overconfidence arising from correlation neglect can lead to stronger partisanship and ideological extremeness as well as increased voter turnout. However, they do not study information aggregation and collective decision making outcomes, which is the focus of our study. Levy and Razin (2015) analyze theoretically information aggregation in elections with correlation neglect and show that correlation neglect may be beneficial for information aggregation by correcting the impact of preference biases on voting behavior. This seemingly contrasts with our results and especially Proposition 3. However, there are several notable differences between Levy and Razin (2015) and the current study. The source of the cognitive bias is different, and individuals are heterogenous in competence in our study. Levy and Razin (2015) allow for a more nuanced utility function than what we present in Proposition 3 which allows for only partisans and neutral individuals. However, abstentions are not allowed in their study. Therefore, our findings complement theirs.

6 Concluding Remarks

In this paper, we show that the joint effect of overconfidence and misinformation can be drastic under certain conditions even if individuals make otherwise fully rational choices. As a result, this drastic impact obtains under more relaxed conditions on overconfidence and misinformation if we take a boundedly rational approach, such as the cursed equilibrium by Eyster and Rabin (2005).

Our paper focuses on the cognitive mechanism behind the influence of misinformation

[42] Correlation neglect is absent in our model as every individual receives only one news signal. However, given that correlation neglect is a cognitive bias, it may be empirically correlated with the Dunning-Kruger effect. In addition, the impact of correlation neglect in our model depends more on the interaction of correlation neglect with q and $p(q)$ (e.g., correlation neglect is especially likely to sway elections involving overconfident correlation neglectors with low q values).

on judgments and opinion formation. Individuals' judgment may also be affected by their ideology or partisanship even when they desire to arrive at an accurate opinion. As a result, there are two possible mechanisms that can explain the influence of misinformation: a motivated reasoning mechanism suggesting that the effect of misinformation is associated with ideology or partisanship, and a cognitive mechanism where it is associated with analytic reasoning and critical thinking skills, which we denote by "competence" in this paper. While we focus on the role of competence in opinion formation, our model generates a (rational) motivated reasoning mechanism in turnout behavior when individuals have different concerns for one type of mistake relative to the other as explained in Section 2.3.

Our model can also be extended to characterize equilibrium behavior allowing for (common knowledge) heterogeneous priors regarding the state of the world due to ideology or partisanship as well as biased/motivated updating from the prior. We conjecture that the results of our paper are robust assuming that only one prior is correct based on the best available evidence and knowledge in the public domain. However, the analysis of this extension is beyond the scope of this paper and a direction for future research. Analyzing the effect of overconfidence on information acquisition and information aggregation is another important avenue for future research.

References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social media and fake news in the 2016 election." *Journal of Economic Perspectives*, 31(2), 211-36.

Austen-Smith, David, and Jeffrey S. Banks. 1996. "Information aggregation, rationality, and the Condorcet jury theorem." *American Political Science Review* 90(1): 34-45.

Bagdikian, B.H. 2004. *The new media monopoly: A completely revised and updated edition with seven new chapters.* Beacon Press.

Barber, B. M., and Odean, T. 2001. "Boys Will be Boys: Gender, Overconfidence, and Common Stock Investment." *Quarterly Journal of Economics*, 116(1), 261-292.

Battaglini, Marco, Rebecca B. Morton, and Thomas R. Palfrey. 2010. "The swing voter's curse in the laboratory." *Review of Economic Studies* 77(1): 61-89.

Benoît, Jean-Pierre, and Juan Dubra. 2011. "Apparent overconfidence." *Econometrica* 79(5): 1591-1625.

Benoît, Jean-Pierre, Juan Dubra, and Don A. Moore. 2015. "Does the better-than-average effect show that people are overconfident?: Two experiments." *Journal of the European Economic Association* 13(2): 293-329.

Bhattacharya, Sourav. 2013. "Preference monotonicity and information aggregation in elections." *Econometrica* 81(3): 1229-1247.

Bhattacharya, Sourav, John Duffy, and Sun-Tak Kim. 2014. "Compulsory versus voluntary voting: An experimental study." *Games and Economic Behavior* 84: 111-131.

Bouton, Laurent, Aniol Llorente-Saguer, and Frédéric Malherbe. 2018. "Get rid of unanimity rule: The superiority of majority rules with veto power." *Journal of Political Economy* 126(1): 107-149.

Camerer, C. and Lovallo, D. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 89(1), 306-318.

Duggan, John, and César Martinelli. 2001. "A Bayesian model of voting in juries." *Games and Economic Behavior*, 37(2): 259-294.

Elbittar, Alexander, Andrei Gomberg, César Martinelli, and Thomas R. Palfrey. 2016. "Ignorance and bias in collective decisions." *Journal of Economic Behavior & Organization.*

Esponda, Ignacio. 2008. "Behavioral equilibrium in economies with adverse selection." *American Economic Review* 98 (4), 1269–1291.

Esponda, Ignacio, and Emanuel Vespa. 2014. "Hypothetical thinking and information extraction in the laboratory." *American Economic Journal: Microeconomics* 6(4): 180-202.

Eyster, Eric and Matthew Rabin. 2005. "Cursed equilibrium." *Econometrica* 73 (5), 1623–1672.

Feddersen, Timothy J., and Wolfgang Pesendorfer. 1996. "The swing voter's curse." *American Economic Review* 86(3): 408-424.

Feddersen, Timothy J., and Wolfgang Pesendorfer. 1999. "Abstention in elections with asymmetric information and diverse preferences." *American Political Science Review* 93(2): 381-398.

Flynn, D. J., Brendan Nyhan, and Jason Reifler. 2017. "The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics." *Political Psychology*, 38: 127-150.

Goeree, Jacob K., and Leeat Yariv. 2011. "An experimental study of collective deliberation." *Econometrica* 79(3): 893-921.

Hermann, E., and Noam Chomsky, N. 2002. Manufacturing Consent: The political economy of the mass media. New York: Pantheon.

Herrera, Helios, Aniol Llorente-Saguer, and Joseph McMurray. 2019a. "The Marginal Voter's Curse." *The Economic Journal* 129(624): 3137-3153.

Herrera, Helios, Aniol Llorente-Saguer, and Joseph McMurray. 2019b. "Information Aggregation and Turnout in Proportional Representation: A Laboratory Experiment." *Journal of Public Economics* 179: 104051.

Jang, S. Mo, and Joon K. Kim. 2018. "Third person effects of fake news: Fake news regulation and media literacy interventions." *Computers in Human Behavior* 80: 295-302.

Kartal, Melis. 2015. "A comparative welfare analysis of electoral systems with endogenous turnout." *The Economic Journal* 125(587): 1369-1392.

Kartal, Melis, and James Tremewan. 2018. "An offer you can refuse: the effect of transparency with endogenous conflict of interest." *Journal of Public Economics,* 161: 44-55.

King Jr, Martin Luther. 1947. "The purpose of education." *The maroon tiger,* 10: 123-124.

Kőszegi, Botond. 2006. "Ego utility, overconfidence, and task choice." *Journal of the European Economic Association* 4(4): 673-707.

Krishna, Vijay, and John Morgan. 2012. "Voluntary voting: Costs and benefits." *Journal of Economic Theory* 147(6): 2083-2123.

Kruger, Justin, and David Dunning. 1999. "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments." *Journal of Personality and Social Psychology* 77(6): 1121.

Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, David Schwieder, and Robert F. Rich. 2000. "Misinformation and the currency of democratic citizenship." *Journal of Politics* 62(3): 790-816.

Kunda, Ziva. 1990. "The case for motivated reasoning." *Psychological bulletin* 108(3): 480.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. 2018. "The science of fake news." *Science*, 359(6380), 1094-1096.

Levy, Gilat, and Ronny Razin. 2015. "Correlation neglect, voting behavior, and information aggregation." *American Economic Review* 105(4): 1634-45.

Lippmann, Walter. 1922. *Public opinion.* Routledge.

Lippmann, Walter, and Charles Merz. 1920. "A Test of the News." A Supplement to the New Republic of August 4th, 1920.

Malmendier, U., and Tate, G. 2005. "CEO Overconfidence and Corporate Investment." *Journal of Finance*, 60(6), 2661–2700.

Marlow, Thomas, Sean Miller, and J. Timmons Roberts. 2020. "Twitter Discourses on Climate Change: Exploring Topics and the Presence of Bots." *Working paper.*

Martinelli, César. 2006. "Would rational voters acquire costly information?." *Journal of Economic Theory* 129(1): 225-251.

McLennan, Andrew. 1998. "Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents." *American Political Science Review* 92(2): 413-418.

McMurray, Joseph C. 2012. "Aggregating information by voting: The wisdom of the experts versus the wisdom of the masses." *Review of Economic Studies* 80(1): 277-312.

Moore, Don A., and Daylian M. Cain. "Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition." *Organizational Behavior and Human Decision Processes* 103(2): 197-213.

Moore, Don A. and Paul J. Healy. 2008. "The trouble with overconfidence." *Psychological Review*, 115(2): 502-517.

Morton, Rebecca B., and Jean-Robert Tyran. 2011. "Let the experts decide? Asymmetric information, abstention, and coordination in standing committees." *Games and Economic Behavior* 72(2): 485-509.

Morton, Rebecca B., Marco Piovesan, and Jean-Robert Tyran. 2019. "The dark side of the vote: Biased voters, social information, and information aggregation through majority voting." *Games and Economic Behavior*, 113: 461-481.

Ortoleva, Pietro, and Erik Snowberg. 2015. "Overconfidence in political behavior." *American Economic Review* 105(2): 504-35.

Pew Research Center, December, 2016. "Many Americans Believe Fake News Is Sowing Confusion"

Taber, Charles S., and Milton Lodge. 2006. "Motivated skepticism in the evaluation of political beliefs." *American Journal of Political Science* 50(3): 755-769.

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

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science* 359(6380): 1146-1151.

A Proofs and Additional Results

A.1 Proofs of Results in Section 2.1

Proof of Lemma 1. The general statement of Lemma 1 is proved in the main text in the discussion preceeding Lemma 1. We now show that in an equilibrium where the correct policy is chosen with a probability (weakly) greater than 0.5 in either state, $q^a = \Pr(S = B|piv_a)$ and $q^b = \Pr(S = A|piv_b)$ must hold. Assume that the correct policy is chosen with a probability (weakly) greater than 0.5 in either state in equilibrium. This is true if and only if the relative turnout rate for the correct policy is weakly greater than 0.5. As a result, $\Pr(piv_b|A) \geq \Pr(piv_a|A)$ and $\Pr(piv_a|B) \geq \Pr(piv_b|B)$ must hold. Now, suppose towards a contradiction that $q^a \neq \Pr(S = B|piv_a)$. This implies that $\Pr(S = B|piv_b) > \Pr(S = B|piv_a)$ by (2) in Lemma 1 statement, which in turn implies that $\Pr(S = A|piv_a) > \Pr(S = A|piv_b)$. Thus, $q^b \neq \Pr(S = A|piv_b)$ by (3) in Lemma 1 statement. Rewriting $\Pr(S = B|piv_b) > \Pr(S = B|piv_a)$ and using $\Pr(piv_b|A) \geq \Pr(piv_a|A)$, we obtain

$$\frac{(1-\pi)\Pr(piv_b|B)}{(1-\pi)\Pr(piv_b|B) + \pi\Pr(piv_a|A)} > \frac{(1-\pi)\Pr(piv_a|B)}{(1-\pi)\Pr(piv_a|B) + \pi\Pr(piv_a|A)},$$

which implies that $\Pr(piv_b|B) > \Pr(piv_a|B)$ in contradiction to the statement above. Finally, as argued above $q^b = \Pr(S = A|piv_b)$ if and only if $q^a = \Pr(S = B|piv_a)$. Hence, the result is proved.

We next show that if $\pi = 0.5$, then $q^a = \Pr(S = B|piv_a)$ and $q^b = \Pr(S = A|piv_b)$. It is enough to show that $q^a = \Pr(S = B|piv_a)$ given the statement above. Assume towards a contradiction that $\Pr(S = B|piv_a) < \Pr(S = B|piv_b)$. Then, given the definitions of q^a and q^b in (2)-(3), it can be checked that there is full turnout (i.e. no abstention) in equilibrium, and $q^a + q^b = 1$ must hold. Thus, either $q^a \geq 0.5$ or $q^b \geq 0.5$ (or $q^a = q^b = 0.5$). Due to symmetry with $\pi = 0.5$, it is without loss of generality to only consider the case in which $q^a \geq 0.5$ and show that there is a contradiction. In particular, we will show that

$$q^a \Pr(piv_a|A) - (1-q^a)\Pr(piv_a|B) = (1-q^a)\Pr(piv_b|B) - q^a \Pr(piv_b|A)$$

cannot hold, but it is a necessary condition as an individual who obtains an α signal and has $q_i = q^a$ must be indifferent between voting for policy a and for policy b. There are two cases to consider: (i) $q^a > 0.5$ and (ii) $q^a = 0.5$. First, assume that $q^a > 0.5$. There are five cases to consider depending on the values of q^a, $\underline{q}$ and $\bar{q}$.

Case 1: $q^a \geq \bar{q}$, and $1 - q^a \leq \underline{q}$. Note that $1 - q^a \leq \underline{q}$ and $q^a \geq \bar{q}$ imply that $q^b \leq \underline{q}$ and $1 - q^b \geq \bar{q}$ by $q^a + q^b = 1$. Thus, this is analogous to a nonresponsive equilibrium in which no individual is pivotal since every individual votes for b regardless of their signal and accuracy, which we rule out.

Case 2: $\underline{q} < q^a < \bar{q}$, and $1 - q^a > \underline{q}$.[43] In this case, the benefit from voting for a for an individual who obtains an α signal with an accuracy of q^a equals

$$q^a \binom{N-1}{\lfloor \frac{N-1}{2} \rfloor} \lambda_A(a)^{\lfloor \frac{N-1}{2} \rfloor} \lambda_A(b)^{N-1-\lfloor \frac{N-1}{2} \rfloor} - (1-q^a) \binom{N-1}{\lfloor \frac{N-1}{2} \rfloor} \lambda_B(a)^{\lfloor \frac{N-1}{2} \rfloor} \lambda_B(b)^{N-1-\lfloor \frac{N-1}{2} \rfloor} \quad (4)$$

where $\lambda_S(a)$ and $\lambda_S(b)$ represent the respective turnout rate for policy a and policy b in state S; i.e.,

$$\begin{aligned}
\lambda_A(a) &= \textstyle\int_{q^a}^{\bar{q}} q dF + \int_{\underline{q}}^{1-q^a} (1-q) dF \\
\lambda_A(b) &= \textstyle\int_{\underline{q}}^{q^a} q dF + \int_{1-q^a}^{\bar{q}} (1-q) dF \\
\lambda_B(a) &= \textstyle\int_{q^a}^{\bar{q}} (1-q) dF + \int_{\underline{q}}^{1-q^a} q dF \\
\lambda_B(b) &= \textstyle\int_{\underline{q}}^{q^a} (1-q) dF + \int_{1-q^a}^{\bar{q}} q dF
\end{aligned}$$

and the same individual's benefit from voting for b equals

$$(1-q^a) \binom{N-1}{\lfloor \frac{N-1}{2} \rfloor} \lambda_B(b)^{\lfloor \frac{N-1}{2} \rfloor} \lambda_B(a)^{N-1-\lfloor \frac{N-1}{2} \rfloor} - q^a \binom{N-1}{\lfloor \frac{N-1}{2} \rfloor} \lambda_A(b)^{\lfloor \frac{N-1}{2} \rfloor} \lambda_A(a)^{N-1-\lfloor \frac{N-1}{2} \rfloor}. \quad (5)$$

Recall that since $q^a + q^b = 1$, there is no abstention and thus $\lambda_S(a) + \lambda_S(b) = 1$ for $S \in \{A, B\}$. However, we will show that (4) is strictly greater than (5), which is a contradition. To see why, note that since $q^a > 0.5$, $\lambda_S(a) + \lambda_S(b) = 1$, and $\lambda_B(b)$ is the largest turnout term, $\lambda_A(a)\lambda_A(b) > \lambda_B(a)\lambda_B(b)$ must hold, and therefore,

$$q^a \binom{N-1}{\lfloor \frac{N-1}{2} \rfloor} \lambda_A(a)^{\lfloor \frac{N-1}{2} \rfloor} \lambda_A(b)^{\lfloor \frac{N-1}{2} \rfloor} (\lambda_A(a) + \lambda_A(b))$$

is greater than

$$(1-q^a) \binom{N-1}{\lfloor \frac{N-1}{2} \rfloor} \lambda_B(b)^{\lfloor \frac{N-1}{2} \rfloor} \lambda_B(a)^{\lfloor \frac{N-1}{2} \rfloor} (\lambda_B(a) + \lambda_B(b))$$

which ensures that benefit from voting for a exceeds the benefit from voting for b with $s_i = \alpha$ and $q_i = q^a$, regardless of whether N is even or odd. Thus, there cannot be an equilibrium with $q^a < \bar{q}$, and $1 - q^a > \underline{q}$.

[43] Note that $1 - q^a \leq \bar{q}$ must always hold by the initial hypothesis that $q^a > 0.5$.

Case 3: $\underline{q} < q^a < \bar{q}$, and $1 - q^a \leq \underline{q}$. The steps in the proof for Case 2 still applies (with minor modifications in $\lambda_S(a)$ and $\lambda_S(b)$ due to $1 - q^a \leq \underline{q}$).
Case 4: $q^a \geq \bar{q}$, and $1 - q^a > \underline{q}$. The steps in the proof for Case 2 still applies (with minor modifications in $\lambda_S(a)$ and $\lambda_S(b)$ due to $q^a \geq \bar{q}$).
Case 5: $q^a \leq \underline{q}$. This case is possible only if $\underline{q} > 0.5$ since $q^a > 0.5$ by initial hypothesis. It also follows from $q^a + q^b = 1$ that $q^b < \underline{q}$ must hold. In this case, everyone votes and does so according to their signal. However, since we have a symmetric environment with $\pi = 0.5$, this implies that $q^a = q^b$. But then from $q^a + q^b = 1$, $q^a = 0.5$ must hold, a contradiction.

Hence, we have shown that $\Pr(S = B|piv_a) < \Pr(S = B|piv_b)$ cannot hold if $q^a > 0.5$ and $\pi = 0.5$. Next, assume that $q^a = 0.5$. From $q^a + q^b = 1$, $q^b = 0.5$ must hold. It is easy to see that this results in a case in which the correct policy is chosen with a probability greater than 0.5 in either state, and therefore, $\Pr(S = B|piv_a) < \Pr(S = B|piv_b)$ cannot hold given what we proved above.

Finally, we show that if $\pi = 0.5$ and $\underline{q} = 0.5$, then $q^a = q^b$. Suppose towards a contradiction (and without loss of generality) that $q^a > q^b$. There are three possibilities: either $q^a > q^b \geq 0.5$ or $q^a > 0.5 > q^b$ or $0.5 \geq q^a > q^b$.

(i) First, assume that $q^a > q^b \geq 0.5$.[44] In that case, voting is informative as no individual votes against his/her signal. In particular, by Lemma 1 every i with $s_i = \alpha$ and $q_i \geq q^a$ votes for policy a and every i with $s_i = \beta$ and $q_i \geq q^b$ votes for policy b. Consider the benefit from voting for a for an individual with $s_i = \alpha$ and $q_i = q^a$. This benefit, which we denote by $\Pi^a(q^a, \alpha)$ equals

$$q^a \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_A)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} q dF\right)^{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^b}^{\bar{q}} (1-q) dF\right)^{t-\lfloor \frac{t}{2} \rfloor} -$$
$$(1-q^a) \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_B)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} (1-q) dF\right)^{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^b}^{\bar{q}} q dF\right)^{t-\lfloor \frac{t}{2} \rfloor}$$

where λ_S represents the turnout rate in state S; i.e., $\lambda_A = \int_{q^a}^{\bar{q}} q dF + \int_{q^b}^{\bar{q}} (1-q) dF$ and $\lambda_B = \int_{q^a}^{\bar{q}} (1-q) dF + \int_{q^b}^{\bar{q}} q dF$. Next, consider the benefit from voting for b for an individual

[44] Note that $q^a > q^b \geq \bar{q}$ can never hold as this implies no one votes in equilibrium, a contradiction. Our proof also works if $q^a > \bar{q}$ as pivotality conditions below adjust to that.

with $s_i = \beta$ and $q_i = q^b$. This benefit, which we denote by $\Pi^b(q^b, \beta)$ equals

$$q^b \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_B)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} (1-q)dF\right)^{t-\lfloor \frac{t}{2} \rfloor} \left(\int_{q^b}^{\bar{q}} qdF\right)^{\lfloor \frac{t}{2} \rfloor} -$$
$$(1-q^b) \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_A)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} qdF\right)^{t-\lfloor \frac{t}{2} \rfloor} \left(\int_{q^b}^{\bar{q}} (1-q)dF\right)^{\lfloor \frac{t}{2} \rfloor}.$$

Note that an individual who obtains an α signal with an accuracy of q^a is indifferent between voting for a and abstaining, and an individual who obtains a β signal with an accuracy of $q^b \geq 0.5$ weakly prefers voting for b over abstaining; i.e., $\Pi^a(q^a, \alpha) = \Pi^b(q^b, \beta) = 0$ must hold. However, we will show that $\Pi^a(q^a, \alpha) > \Pi^b(q^b, \beta)$ resulting in a contradiction. To see why, first note that $\lambda_A < \lambda_B$ and thus, $1-\lambda_A > 1-\lambda_B$. Then, to show that $\Pi^a(q^a, \alpha) > \Pi^b(q^b, \beta)$, first note that for realized turnout $t = 0$,[45] $q^a + 1 - q^b > q^b + 1 - q^a$, for $t > 0$ even,

$$\int_{q^a}^{\bar{q}} qdF \int_{0.5}^{q^b} (1-q)dF > \int_{q^a}^{\bar{q}} (1-q)dF \int_{q^b}^{\bar{q}} qdF \tag{6}$$

and for $t > 0$ odd

$$q^a \int_{q^b}^{\bar{q}} (1-q)dF + (1-q^b) \int_{q^a}^{\bar{q}} qdF > q^b \int_{q^a}^{\bar{q}} (1-q)dF + (1-q^a) \int_{q^b}^{\bar{q}} qdF. \tag{7}$$

It can easily be checked that (6) holds as $\int_{q^a}^{\bar{q}} qdF/(1-F(q^a)) > \int_{0.5}^{q^b} qdF/(1-F(q^b))$. Suppose towards a contradiction that (7) does not hold. But this implies that

$$q^a(1-F(q^b)) - q^b(1-F(q^a)) \leq \int_{q^b}^{q^a} qdF,$$

which cannot hold as $q^a > q^b$ and $q^a(F(q^a) - F(q^b)) > \int_{q^b}^{q^a} qdF$.

(ii) Next, assume that $q^a \geq 0.5 > q^b$. In that case, voting is informative only for those who obtain a β signal. In particular, since $0.5 > q^b$, by Lemma 1 every individual who obtains a β signal votes for policy b, every i who obtains an α signal but has $q_i \leq 1 - q^b$ votes for policy b and finally, every i who obtains an α signal and has $q_i \geq q^a$ votes for policy a (recall that from the definition of q^a and q^b, we have $q^a \geq 1 - q^b$). Consider the benefit from voting for a for an individual with $s_i = \alpha$ and $q_i = q^a$. This benefit, which we denote by $\Pi^a(q^a, \alpha)$ equals

$$q^a \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_A)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} qdF\right)^{\lfloor \frac{t}{2} \rfloor} \left(\int_{0.5}^{1-q^b} qdF + \int_{0.5}^{\bar{q}} (1-q)dF\right)^{t-\lfloor \frac{t}{2} \rfloor} -$$
$$(1-q^a) \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_B)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} (1-q)dF\right)^{\lfloor \frac{t}{2} \rfloor} \left(\int_{0.5}^{1-q^b} (1-q)dF + \int_{0.5}^{\bar{q}} qdF\right)^{t-\lfloor \frac{t}{2} \rfloor}$$

[45] Note that if $q^a > \bar{q}$, only $t = 0$ and $t = 1$ are relevant.

where λ_S represents the turnout rate in state S; i.e., $\lambda_A = \int_{q^a}^{\bar{q}} qdF + \int_{0.5}^{1-q^b} qdF + \int_{0.5}^{\bar{q}} (1-q)dF$, and $\lambda_B = \int_{q^a}^{\bar{q}} (1-q)dF + \int_{0.5}^{1-q^b} (1-q)dF + \int_{0.5}^{\bar{q}} qdF$. Next, consider the benefit from voting for b for an individual with $s_i = \beta$ and $q_i = 1 - q^b$ (note that by hypothesis $1 - q^b > 0.5$).[46] This benefit, which we denote by $\Pi^b(1-q^b, \alpha)$ equals

$$q^b \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_B)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} (1-q)dF \right)^{t - \lfloor \frac{t}{2} \rfloor} \left(\int_{0.5}^{1-q^b} (1-q)dF + \int_{0.5}^{\bar{q}} qdF \right)^{\lfloor \frac{t}{2} \rfloor} -$$
$$(1-q^b) \sum_{t=0}^{N-1} \binom{N-1}{t} (1-\lambda_A)^{N-1-t} \binom{t}{\lfloor \frac{t}{2} \rfloor} \left(\int_{q^a}^{\bar{q}} qdF \right)^{t - \lfloor \frac{t}{2} \rfloor} \left(\int_{0.5}^{1-q^b} qdF + \int_{0.5}^{\bar{q}} (1-q)dF \right)^{\lfloor \frac{t}{2} \rfloor}.$$

First, assume that $q^a > 1-q^b$. Then, an individual who obtains an α signal with an accuracy of q^a is indifferent between voting for a and abstaining, and an individual who obtains an α signal with an accuracy of $1 - q^b > 0.5$ is indifferent between voting for b and abstaining; i.e., $\Pi^a(q^a, \alpha) = \Pi^b(1-q^b, \alpha) = 0$ must hold. Next, assume that $q^a = 1 - q^b$. In this case, there is no abstention in equilibrium and an individual who obtains an α signal with an accuracy of q^a is indifferent between voting for a and voting for b (and prefers either one over abstaining). In equilibrium, $\Pi^a(q^a, \alpha) = \Pi^b(1-q^b, \alpha)$ must hold. However, we will now show that $\Pi^a(q^a, \alpha) > \Pi^b(1-q^b, \alpha)$ must hold, which results in a contradiction. To see why, first note that $\lambda_A < \lambda_B$ and thus, $1 - \lambda_A > 1 - \lambda_B$ if $q^a > 1 - q^b$ and $\lambda_A = \lambda_B = 1$ if $q^a = 1 - q^b$. Then, to show $\Pi^a(q^a, \alpha) > \Pi^b(1-q^b, \alpha)$, first note that for realized turnout $t = 0$, it is enough to see that $q^a + 1 - q^b > q^b + 1 - q^a$ and for $t > 0$, it is enough to show that

$$\left(\int_{q^a}^{\bar{q}} qdF \right) \left(\int_{0.5}^{1-q^b} qdF + \int_{0.5}^{\bar{q}} (1-q)dF \right) > \left(\int_{q^a}^{\bar{q}} (1-q)dF \right) \left(\int_{0.5}^{1-q^b} (1-q)dF + \int_{0.5}^{\bar{q}} qdF \right) \quad (8)$$

and that

$$q^a \left(\int_{0.5}^{1-q^b} qdF + \int_{0.5}^{\bar{q}} (1-q)dF \right) + \left(1-q^b\right) \int_{q^a}^{\bar{q}} qdF > \quad (9)$$
$$q^b \int_{q^a}^{\bar{q}} (1-q)dF + (1-q^a) \left(\int_{0.5}^{1-q^b} (1-q)dF + \int_{0.5}^{\bar{q}} qdF \right)$$

Suppose towards a contradiction that (8) does not hold. This implies that

$$\int_{q^a}^{\bar{q}} qdF \leq \int_{q^a}^{\bar{q}} dF \int_{0.5}^{1-q^b} dF - \int_{q^a}^{\bar{q}} qdF \int_{0.5}^{1-q^b} dF - \int_{q^a}^{\bar{q}} dF \int_{0.5}^{1-q^b} qdF + \int_{q^a}^{\bar{q}} dF \int_{0.5}^{\bar{q}} qdF$$

must hold. This inequality implies in turn that

$$\int_{q^a}^{\bar{q}} qdF(1 + F(1-q^b)) + (1 - F(q^a) \int_{0.5}^{1-q^b} qdF \leq (1 - F(q^a))(F(1-q^b) + \int_{0.5}^{\bar{q}} qdF).$$

[46] We assume that $1 - q^b \leq \bar{q}$ which is without loss of generality for the results.

Dividing both sides by $1-F(q^a)$ and then taking $\int_{0.5}^{1-q^b} qdF$ to the right-hand side, we obtain

$$\frac{\int_{q^a}^{\bar{q}} qdF}{1-F(q^a)}(1+F(1-q^b)) \leq F(1-q^b) + \int_{1-q^b}^{\bar{q}} qdF.$$

However, writing $(1+F(1-q^b))$ above as $(1-F(1-q^b)+2F(1-q^b))$ and noting that $\frac{\int_{q^a}^{\bar{q}} qdF}{1-F(q^a)} > q^a$, it can be checked that the right-hand side of the inequality above is strictly greater than $\frac{\int_{q^a}^{\bar{q}} qdF}{1-F(q^a)}(1-F(1-q^b)) + 2q^aF(1-q^b)$, which in turn is greater than $F(1-q^b) + \int_{1-q^b}^{\bar{q}} qdF$ because $2q^a \geq 1$ and by $q^a \geq 1-q^b$, $\frac{\int_{q^a}^{\bar{q}} qdF}{1-F(q^a)} \geq \frac{\int_{1-q^b}^{\bar{q}} qdF}{1-F(1-q^b)}$ holds, which is a contradiction. Thus, (8) must hold. Next, suppose towards a contradiction that (9) does not hold. This implies that $q^a + \int_{q^a}^{\bar{q}} qdF \leq q^b(1-F(q^a)) + \int_{1-q^b}^{\bar{q}} qdF + (1-q^a)F(1-q^b)$, and so, $q^a(1+F(1-q^b)) \leq q^b(1-F(q^a)) + \int_{1-q^b}^{q^a} qdF + F(1-q^b)$. Noting that $\int_{1-q^b}^{q^a} qdF < q^a(F(q^a)-F(1-q^b))$, this implies $q^a(1+2F(1-q^b)-F(q^a)) < q^b(1-F(q^a)) + F(1-q^b)$. However, this cannot hold as $q^a > q^b$ and $2q^a > 1$.

(iii) Finally, we rule out the case where $0.5 \geq q^a > q^b$. As mentioned above, it can be checked from the definition of q^a and q^b that $q^a \geq 1-q^b$ must hold. Thus, it is not possible to have $0.5 \geq q^a > q^b$.

Proof of Lemma 2. First, consider the case in which $\underline{q} < 0.5$. Let $\hat{q}$ be such that $\hat{q} = \int_{\underline{q}}^{0.5}(1-q)dF + \int_{0.5}^{\bar{q}} qdF$ — this is the expected accuracy if i votes for (against) s_i for every $q_i > 0.5$ ($q_i < 0.5$). Obviously, $\hat{q} > 0.5$. Next, let π^* be such that

$$\pi^* = \sum_{t=\frac{N}{2}+1}^{N} \binom{N}{t}(\hat{q})^t(1-\hat{q})^{N-t} + 0.5\binom{N}{\frac{N}{2}}(\hat{q})^{\frac{N}{2}}(1-\hat{q})^{\frac{N}{2}}$$

for N even, and for N odd, let $\pi^* = \sum_{t=\frac{N+1}{2}}^{N} \binom{N}{t}(\hat{q})^t(1-\hat{q})^{N-t}$. Since $\hat{q} > 0.5$, $\pi^* > 0.5$. In fact, from Lemma 4 below, it follows that $\pi^* > \hat{q}$ for $N \geq 3$ (since $\pi^* = \hat{q}$ if $N = 1$ and $N = 2$). By construction, the optimal symmetric equilibrium must be a responsive equilibrium for every $\pi \in (1-\pi^*\pi^*)$. This is because (i) the efficiency of a nonresponsive equilibrium cannot be greater than $\min\{1-\pi, \pi\}$, (ii) there exists an optimal symmetric strategy by Lemma 3 below, which in turn is the optimal symmetric equilibrium, and by construction, (iii) the expected accuracy in the optimal symmetric equilibrium with $\pi \in (1-\pi^*\pi^*)$ must be higher than π^*. Next, we show that the optimal equilibrium must have an interior cutoff for every $\pi \in (1-\pi^*\pi^*)$. Suppose not. We can immediately rule out the case in which no individual votes. Moreover, the case in which every i who receives an α signal (a β signal) votes for

a (b) and every i who receives a β signal (an α signal) abstains cannot be an equilibrium. It is enough to consider the case where every i who receives an α signal votes for a and every i who receives a β signal abstains. This implies by Lemma 1 that $q^a \leq \underline{q} < 0.5$ (this is necessary for i with α signal and $q_i \geq \underline{q}$ to prefer voting for a over abstention or voting for b), and as a result, $1 - q^a > 0.5$. Thus, every individual with $s_i = \beta$ and $q_i < 1 - q^a$ must strictly prefer voting for a in contradiction to the strategy. In a similar vein, Lemma 1 implications rule out the case in which every i who receives an α signal (a β signal) votes for b (a) and every i who receives a β signal (an α signal) abstains cannot be an equilibrium. Remaining cases are the cases in which every i votes either always for or always against s_i regardless of q_i. Consider the former case. Again, by Lemma 1, this cannot be an equilibrium because if $q^a \leq \underline{q} < 0.5$ then $1 - q^a > 0.5$, and every individual with $s_i = \beta$ and $q_i < 1 - q^a$ must strictly prefer voting for a over abstention or voting for b, a contradiction. The latter strategy can also not be part of an equilibrium as by Lemma 1, $\bar{q} \leq 1 - q^b$ must hold (this is necessary for every i with α signal to prefer voting for b over abstention or voting for a), but this implies that $q^b < 0.5$, and therefore, an individual with a β signal and $q_i > q^b$ (for example $q_i = 0.5$) must strictly prefer voting for b over abstention or voting for a as well, which is a contradiction. Hence, the optimal equilibrium must have an interior cutoff for every $\pi \in (1 - \pi^* \pi^*)$.

Next, consider the case in which $\underline{q} = 0.5$. Let π^* be such that

$$\pi^* = \sum_{t=\frac{N}{2}+1}^{N} \binom{N}{t} (\mathbb{E}(q))^t (1 - \mathbb{E}(q))^{N-t} + 0.5 \binom{N}{\frac{N}{2}} (\mathbb{E}(q))^{\frac{N}{2}} (1 - \mathbb{E}(q))^{\frac{N}{2}}$$

for N even, and for N odd, let $\pi^* = \sum_{t=\frac{N+1}{2}}^{N} \binom{N}{t} (\mathbb{E}(q))^t (1 - \mathbb{E}(q))^{N-t}$, where $\mathbb{E}(q) = \int_{0.5}^{\bar{q}} q dF$. By Lemma 4 proved later, $\pi^* > \mathbb{E}(q)$ if $N \geq 3$ (and $\pi^* \geq \mathbb{E}(q)$ if $N = 2$) First, consider the case in which N is even. As discussed above, for every $\pi \in (1 - \pi^*, \pi^*)$, the optimal equilibrium must be a responsive equilibrium. Next, we show that the optimal equilibrium must have an interior cutoff for $\pi \in (1 - \pi^*, \pi^*)$. Suppose not. We can immediately rule out the case where no individual votes. Next, it can be shown that the case in which every i who receives an α signal (a β signal) votes for a (b) and every i who receives a β signal (an α signal) abstains cannot be an equilibrium (except possibly in one knife-edge case). It is enough to consider the case where every i who receives an α signal votes for a and every i who receives a β signal abstains. Note that in this case $0.5 = \frac{(1-\pi) \Pr(piv_a | S=B)}{\pi \Pr(piv_a | S=A) + (1-\pi) \Pr(piv_a | S=B)}$ and $\bar{q} < 1$ must hold by Lemma 1. This equality requires π to be exactly equal to $\frac{\left(\int_{0.5}^{\bar{q}} q dF\right)^{N-1}}{\left(\int_{0.5}^{\bar{q}} q dF\right)^{N-1} + \left(\int_{0.5}^{\bar{q}} (1-q) dF\right)^{N-1}}$ as well as $\pi > \bar{q}$, which is a nongeneric case we rule out. In a similar vein, Lemma 1 implications rule out the case where every i who receives an α signal (a β signal) votes for b (a) and every i who receives a β signal (an α signal) abstains. Remaining cases are the cases in which every i votes either always for or always against s_i regardless of q_i. Consider the former case. This cannot be an equilibrium because it can easily be shown that given the described strategy, $q^a = q^b > 0.5$ if $\pi = 0.5$ (recall that N is even), $q^a > 0.5$ if $\pi < 0.5$ and $q^b > 0.5$ if $\pi > 0.5$, but this is a contradiction with the strategy since $\underline{q} = 0.5$. Now, consider the latter case. In this case, $\bar{q} \leq 1 - q^b$ must hold by Lemma 1 (this is necessary for i with α signal and $q_i \leq \bar{q}$ to prefer voting for b over voting for a), but this implies that $q^b < 0.5$, and therefore, an individual with a β signal and $q_i > q^b$ (for example, $q_i = 0.5$) must strictly prefer voting for b over abstention or voting for a, which is a contradiction. The proof for the case in which N is odd is analogous except that with $\pi = \frac{1}{2}$ the optimal equilibrium need not involve interior cutoffs.

Finally, consider the case in which $\underline{q} > 0.5$. First, let $\hat{\pi}$ be analogous to π^* as defined in the case with $\underline{q} = 0.5$ above. As shown in Lemma 4, $\hat{\pi} > \mathbb{E}(q)$ if $N \geq 3$ and $\hat{\pi} = \mathbb{E}(q)$ otherwise. We now define $\pi^* = \min\{\bar{q}, \hat{\pi}\}$. First, consider the case in which N is even. Similar to our discussion above with $\underline{q} \leq 0.5$, for every $\pi \in (1 - \pi^*, \pi^*)$, the optimal equilibrium must be a responsive equilibrium (note that $\underline{q} < \mathbb{E}(q) < \pi^*$). We will now show that the optimal equilibrium must have an interior cutoff if $\pi \in (1 - \pi^*, 1 - \underline{q}] \cup [\underline{q}, \pi^*)$. Suppose not. We can immediately rule out the case in which no individual votes. Moreover, by construction, the case in which every i who receives an α signal (a β signal) votes for a (b) and every i who receives a β signal (an α signal) abstains cannot be an equilibrium. It is enough to consider the case where every i who receives an α signal votes for a and every i who receives a β signal abstains. Since $\bar{q} \geq \pi^*$, and since $\Pr(piv_b|S = B) < \Pr(piv_b|S = A)$ given the prescribed strategy, $q^b = \frac{\pi \Pr(piv_b|S=A)}{\pi \Pr(piv_b|S=A)+(1-\pi)\Pr(piv_b|S=B)} < \pi < \bar{q}$ holds for every $\pi \in (1-\pi^*, 1-\underline{q}] \cup [\underline{q}, \pi^*)$. Thus, abstention is not optimal for i if $s_i = \beta$ and q_i is sufficiently high. As in the previous cases above with $\underline{q} \leq 0.5$, Lemma 1 implications rule out the case where every i who receives an α signal (a β signal) votes for b (a) and every i who receives a β signal (an α signal) abstains. Remaining cases are the cases in which every i votes either always for or always against s_i regardless of q_i. Consider the former case. This cannot be an equilibrium because given the described strategy, $q^a > \underline{q}$ must hold if $\pi \leq 1 - \underline{q}$ (because $\Pr(piv_a|S = A) < \Pr(piv_a|S = B)$ and $q^a = \frac{(1-\pi)\Pr(piv_a|S=B)}{\pi \Pr(piv_a|S=A)+(1-\pi)\Pr(piv_a|S=B)} > 1 - \pi \geq \underline{q}$) and similarly $q^b > \underline{q}$ must hold if $\pi \geq \underline{q}$, which contradicts the voting strategy described. Now, consider the latter case. In this case, $\bar{q} \leq 1 - q^b$ must hold by Lemma 1, but this implies that $q^b < 0.5$, and therefore, an individual with a β signal and $q_i > q^b$ (for example, $q_i = 0.5$) must strictly prefer voting for b over abstention or voting for a, which is a contradiction. The proof for the case where N is odd and $\pi \in (1-\pi^*, 1-\underline{q}) \cup (\underline{q}, \pi^*)$ is analogous.

Proof of Proposition 1. We first prove the existence of the optimal symmetric strategy in Lemma 3. To do that, we first define a "cutoff strategy": a cutoff strategy consists of four cutoffs $\underline{q}^j \in [\underline{q}, \bar{q}]$ and $\bar{q}^j \in [\underline{q}, \bar{q}]$, $j \in \{a, b\}$ such that individual i votes for policy a if $s_i = \alpha$ and $q_i \geq \bar{q}^a$ or if $s_i = \beta$ and $q_i \leq \underline{q}^a$, and i votes for policy b if $s_i = \beta$ and $q_i \geq \bar{q}^b$ or if $s_i = \alpha$ and $q_i \leq \underline{q}^b$.

Lemma 3: For every symmetric (measurable-)strategy that is not a cutoff strategy, there exists a cutoff strategy that strictly dominates it. As a result, by the Weierstrass theorem, there exists an optimal strategy among all symmetric measurable strategies.

Proof: Let $\lambda : [\underline{q}, \bar{q}] \times \{\alpha, \beta\} \to [0,1] \times [0,1]$ represent a strategy that maps every q_i and s_i to a probability of voting for the policy that matches s_i and to a probability of voting for the opposite policy (i abstains with the remaining probability). Next, let $T_j(S)$ denote the expected turnout rate for policy $j \in \{a, b\}$ in state $S \in \{A, B\}$. Fix an arbitrary symmetric strategy that doesn't have the cutoff form. In particular, abusing notation let $\lambda_s^j(q)$ represent the probability with which i votes for $j \in \{a, b\}$ given $s_i = s$ and $q_i = q$, and consider a strategy such that $\lambda_s^j(q)$ does not have a cutoff form for at least one (j, s) pair. Assuming that s is the signal consistent with state S (i.e., α for A and β for B) and $s \neq s'$,

$$T_j(S) = \int_{\underline{q}}^{\bar{q}} \left(\lambda_s^j(q)q + \lambda_{s'}^j(q)(1-q)\right) dF$$

for policy $j \in \{a, b\}$ in state $S \in \{A, B\}$. We first look for $\underline{q}^a \in [\underline{q}, \bar{q}]$ and $\bar{q}^a \in [\underline{q}, \bar{q}]$ such that $T_a(B)$ remains constant as follows: $\int_{\underline{q}}^{\underline{q}^a} q dF = \int_{\underline{q}}^{\bar{q}} \lambda_\beta^a(q) q dF$ and $\int_{\bar{q}^a}^{\bar{q}} (1-q) dF = \int_{\underline{q}}^{\bar{q}} \lambda_\alpha^a(q)(1-q) dF$. Note that the former implies that

$$\int_{\underline{q}}^{\underline{q}^a} (1-q) dF \geq \int_{\underline{q}}^{\bar{q}} \lambda_\beta^a(q)(1-q) dF$$

(with strict inequality unless $\lambda_\beta^a(q) = \mathbf{1}_{q \leq x}$ for some $x \in [\underline{q}, \bar{q}]$) and the latter implies that

$$\int_{\bar{q}^a}^{\bar{q}} q dF \geq \int_{\underline{q}}^{\bar{q}} \lambda_\alpha^a(q) q dF,$$

(with strict inequality unless $\lambda_\beta^a(q) = \mathbf{1}_{q \geq x}$ for some $x \in [\underline{q}, \bar{q}]$). We first prove the former. Let $\int_{\underline{q}}^{\underline{q}^a} q dF = \int_{\underline{q}}^{\bar{q}} \lambda_\beta^a(q) q dF$. The result is trivial if $\lambda_\beta^a(q) = \mathbf{1}_{q \leq x}$ for some $x \in [\underline{q}, \bar{q}]$, so assume that $\lambda_\beta^a(q) \neq \mathbf{1}_{q \leq x}$ for any $x \in [\underline{q}, \bar{q}]$. Thus, $\underline{q}^a \in (\underline{q}, \bar{q})$. Suppose towards a contradiction that

$\int_{\underline{q}}^{\underline{q}^a}(1-q)dF \le \int_{\underline{q}}^{\bar{q}}\lambda_\beta^a(q)(1-q)dF$. This implies that

$$\int_{\underline{q}}^{\underline{q}^a}dF \le \int_{\underline{q}}^{\bar{q}}\lambda_\beta^a(q)dF = \int_{\underline{q}}^{\underline{q}^a}\lambda_\beta^a(q)dF + \int_{\underline{q}^a}^{\bar{q}}\lambda_\beta^a(q)dF,$$

and thus $\int_{\underline{q}}^{\underline{q}^a}(1-\lambda_\beta^a(q))dF \le \int_{\underline{q}^a}^{\bar{q}}\lambda_\beta^a(q)dF$. Multiplying both sides of the last inequality by $\underline{q}^a \in (\underline{q}, \bar{q})$, the right hand side is strictly lower than $\int_{\underline{q}^a}^{\bar{q}}\lambda_\beta^a(q)qdF$ (since $\underline{q}^a < \bar{q}$)), and the left hand side is strictly greater than $\int_{\underline{q}}^{\underline{q}^a}(1-\lambda_\beta^a(q))qdF$ (since $\underline{q}^a > \underline{q}$)) resulting in a contradiction to the equality $\int_{\underline{q}}^{\underline{q}^a}qdF = \int_{\underline{q}}^{\bar{q}}\lambda_\beta^a(q)qdF$. Next, let $\int_{\bar{q}^a}^{\bar{q}}(1-q)dF = \int_{\underline{q}}^{\bar{q}}\lambda_\alpha^a(q)(1-q)dF$, and suppose towards a contradiction that $\int_{\bar{q}^a}^{\bar{q}}qdF \le \int_{\underline{q}}^{\bar{q}}\lambda_\alpha^a(q)qdF$ where $\lambda_\alpha^a(q) \neq \mathbf{1}_{q\ge x}$ for any $x \in [\underline{q}, \bar{q}]$. These imply that

$$\int_{\bar{q}^a}^{\bar{q}}dF \le \int_{\underline{q}}^{\bar{q}}\lambda_\alpha^a(q)dF = \int_{\underline{q}}^{\bar{q}^a}\lambda_\alpha^a(q)dF + \int_{\bar{q}^a}^{\bar{q}}\lambda_\alpha^a(q)dF,$$

and thus $\int_{\bar{q}^a}^{\bar{q}}(1-\lambda_\alpha^a(q))dF \le \int_{\underline{q}}^{\bar{q}^a}\lambda_\alpha^a(q)dF$. Multiplying both sides of the inequality by $1-\bar{q}^a$, and rearranging we obtain a contradiction to the equality $\int_{\bar{q}^a}^{\bar{q}}(1-q)dF = \int_{\underline{q}}^{\bar{q}}\lambda_\alpha^a(q)(1-q)dF$. As a result, in our construction $T_a(A)$ strictly increases if $\lambda_\beta^a(q) \neq \mathbf{1}_{q\le x}$ for any $x \in [\underline{q}, \bar{q}]$ or if $\lambda_\alpha^a(q) \neq \mathbf{1}_{q\ge x}$ for any $x \in [\underline{q}, \bar{q}]$, and remains constant otherwise.

Next, we look for $\underline{q}^b \in [\underline{q}, \bar{q}^a)$ such that $\int_{\underline{q}}^{\underline{q}^b}qdF = \int_{\underline{q}}^{\bar{q}}\lambda_\alpha^b(q)qdF$. If this equality cannot be satisfied for any $\underline{q}^b \in [\underline{q}, \bar{q}^a)$, then we set $\underline{q}^b = \bar{q}^a$ (in that case, there is no abstention after an α signal). Similarly, we look for $\bar{q}^b \in (\underline{q}^a, \bar{q}]$ such that $\int_{\bar{q}^b}^{\bar{q}}(1-q)dF = \int_{\underline{q}}^{\bar{q}}\lambda_\beta^b(q)(1-q)dF$, and if there exists no $\bar{q}^b \in (\underline{q}^a, \bar{q}]$ that satisfies the equality, then we set $\underline{q}^a = \bar{q}^b$ (in that case, there is no abstention after a β signal). Thus, $T_b(B)$ increases in our construction (strictly if $\lambda_\alpha^b(q) \neq \mathbf{1}_{q\le x}$ for any $x \in [\underline{q}, \bar{q}]$ or if $\lambda_\beta^b(q) \neq \mathbf{1}_{q\ge x}$ for any $x \in [\underline{q}, \bar{q}]$). While $T_b(A)$ may decrease due to having to set $\underline{q}^b = \bar{q}^a$ or $\underline{q}^a = \bar{q}^b$, $T_a(A) + T_b(A)$ cannot decrease in our construction since $\underline{q}^b = \bar{q}^a$ implies that there is no abstention after an α signal and $\underline{q}^a = \bar{q}^b$ implies that there is no abstention after a β signal. Thus, given these four cutoffs we construct from the initial strategy, expected turnout rate $T_a(S) + T_b(S)$ weakly increases in either state, and it can be checked that the "relative turnout rate" for the correct policy weakly increases in both states; i.e., both $\frac{T_a(A)}{T_a(A)+T_b(A)}$ and $\frac{T_b(B)}{T_b(B)+T_a(B)}$ increase. In fact, at least one of these relative turnout rates must strictly increase by construction since $\lambda_s^j(q)$ does not have a cutoff form for at least one (j, s) pair.

To see why, first assume that $T_a(A)$ remains constant in our construction because the initial strategy is such that $\lambda_\beta^a(q) = \mathbf{1}_{q\le \underline{q}^a}$ for $\underline{q}^a \in [\underline{q}, \bar{q}]$ and if $\lambda_\alpha^a(q) = \mathbf{1}_{q\ge \bar{q}^a}$ for $\bar{q}^a \in [\underline{q}, \bar{q}]$ (recall that $T_a(B)$ is always constant by construction). Then, it can be checked that $T_b(A)$ must be constant as well given our construction: there must exist $\underline{q}^b \in [\underline{q}, \bar{q}^a]$ such that

$\int_{\underline{q}}^{\underline{q}^b} qdF = \int_{\underline{q}}^{\bar{q}} \lambda_\alpha^b(q) qdF$, and $\bar{q}^b \in [\underline{q}^a, \bar{q}]$ such that $\int_{\bar{q}^b}^{\bar{q}}(1-q)dF = \int_{\underline{q}}^{\bar{q}} \lambda_\beta^b(q)(1-q)dF$ (otherwise, $T_a(A) + T_b(A) > 1$, a contradiction). If $\underline{q}^b < \bar{q}^a$ or if $\bar{q}^b > \underline{q}^a$, then $T_b(B)$ and $\frac{T_b(B)}{T_b(B)+T_a(B)}$ strictly increase as desired. Indeed one of the two ($\underline{q}^b < \bar{q}^a$ or $\bar{q}^b > \underline{q}^a$) must hold as otherwise $\lambda_s^j(q)$ must be a cutoff strategy in contrast to our initial hypothesis. As a result, if $T_a(A)$ remains constant in our construction, then expected turnout rate $T_a(A)+T_b(A)$ is constant, but $\frac{T_b(B)}{T_b(B)+T_a(B)}$ and $T_a(B) + T_b(B)$ are strictly higher. Next, assume that $T_a(A)$ strictly increases because $\lambda_\beta^a(q) \neq \mathbf{1}_{q\leq x}$ for any $x \in [\underline{q}, \bar{q}]$ and/or $\lambda_\alpha^a(q) \neq \mathbf{1}_{q\geq x}$ for any $x \in [\underline{q}, \bar{q}]$. As discussed above, by construction, $T_a(A)+T_b(A)$ cannot decrease, and $T_b(A)$ weakly decreases. As a result, and given that $T_a(A)$ is strictly higher, $\frac{T_a(A)}{T_a(A)+T_b(A)}$ strictly increases. Moreover, as discussed above, $T_b(B)$ and $T_a(B) + T_b(B)$ weakly increase by construction.

To complete the proof for showing that the cutoff strategy we constructed is strictly better than the non-cutoff strategy $\lambda_s^j(q)$, Lemma 4 and Lemma 5 are sufficient.

Lemma 4: Let $W_t(q)$ be such that

$$W_t(q) = \begin{cases} \sum\limits_{i=\frac{t+1}{2}}^{t} \binom{t}{i} q^i(1-q)^{t-i} & \text{if } t \text{ is odd} \\ \sum\limits_{i=\frac{t}{2}+1}^{t} \binom{t}{i} q^i(1-q)^{t-i} + \frac{1}{2}\binom{t}{\frac{t}{2}} q^{\frac{t}{2}}(1-q)^{\frac{t}{2}} & \text{if } t \text{ is even} \end{cases}$$

where $q > 1/2$. Then, $W_t(q)$ is monotonic in t. In particular, $W_t(q) = W_{t+1}(q) < W_{t+2}(q)$ for every odd $t > 0$. Moreover, $W_1(q) = W_2(q) = q$ and, thus, $W_t(q) > q$ for every $t > 2$.

Proof: First, we show that $W_t(q) = W_{t+1}(q)$ if t is odd. Let $\Pr(X \leq k; t, q)$ denote the cumulative Binomial distribution function with t trials, X successes and success probability q; i.e.,

$$\Pr(X \leq k; t, q) = (t-k)\binom{t}{k} \int_0^{1-q} x^{t-k-1}(1-x)^k dx$$

Using this formula, it can be checked that $\Pr(X \leq \frac{t+1}{2}; t+1, q) - \frac{1}{2}\Pr(X = \frac{t+1}{2}; t+1, q) = \Pr(X \leq \frac{t-1}{2}; t, q)$ must hold.

Next, we show that $W_t(q) < W_{t+2}(q)$ if t is odd. First, we prove that $W_{t+2}(q) > W_t(q)$ for every $q \geq q^* > \frac{1}{2}$, where q^* is given by $q^*(1-q^*) = \frac{1}{4}\frac{t+1}{t+2}$. Then, for $t > 0$ odd, $W_{t+2}(q) - W_t(q)$ equals

$$\frac{t+1}{2}\binom{t}{\frac{t-1}{2}} \int_0^{1-q} x^{\frac{t-1}{2}}(1-x)^{\frac{t-1}{2}} dt - \frac{t+3}{2}\binom{t+2}{\frac{t+1}{2}} \int_0^{1-q} x^{\frac{t+1}{2}}(1-x)^{\frac{t+1}{2}} dx$$

which equals

$$\int_0^{1-q} x^{\frac{t-1}{2}}(1-x)^{\frac{t-1}{2}} \left(\frac{t+1}{2}\binom{t}{\frac{t-1}{2}} - \frac{t+3}{2}\binom{t+2}{\frac{t+1}{2}} x(1-x)\right) dx.$$

It can be checked that this term is strictly greater than 0 for all $q \geq q^* > \frac{1}{2}$, where $q^* > \frac{1}{2}$ is given by $q^*(1-q^*) = \frac{1}{4}\frac{t+1}{t+2}$. We now show that $W_t(\frac{1}{2}) = W_{t+2}(\frac{1}{2})$ and that $\frac{d}{dq}\left(W_{t+2}(q) - W_t(q)\right) > 0$ for every $q \in [\frac{1}{2}, q^*)$, where q^* is given by $q^*(1-q^*) = \frac{1}{4}\frac{t+1}{t+2}$, which will imply that $W_{t+2}(q) > W_t(q)$ for every $q \in (\frac{1}{2}, q^*)$ and complete the proof. It can be checked that $\frac{d}{dq}\left(W_{t+2}(q) - W_t(q)\right)$ is equal to $q^{\frac{t-1}{2}}(1-q)^{\frac{t-1}{2}}\left(\frac{t+3}{2}\binom{t+2}{\frac{t+1}{2}}q(1-q) - \frac{t+1}{2}\binom{t}{\frac{t-1}{2}}\right)$, which is strictly positive for all $q \in [\frac{1}{2}, q^*)$. Moreover, $W_t(\frac{1}{2}) = W_{t+2}(\frac{1}{2})$ because it can be checked that $\sum\limits_{i=\frac{t+1}{2}}^{t}\binom{t}{i}\left(\frac{1}{2}\right)^t = \sum\limits_{i=\frac{t+3}{2}}^{t+2}\binom{t+2}{i}\left(\frac{1}{2}\right)^{t+2}$ using the fact that $\sum\limits_{i=0}^{t}\binom{t}{i} = 2\sum\limits_{i=\frac{t+1}{2}}^{t}\binom{t}{i} = 2^t$.

Lemma 5: *If p and q increase to $p' \geq p$ and $q' \geq q$ respectively (with at least one strict inequality), then*

$$\sum_{t=0}^{N}\binom{N}{t}p^t(1-p)^{N-t}W_t(q) < \sum_{t=0}^{N}\binom{N}{t}(p')^t(1-p')^{N-t}W_t(q').$$

Proof: First, we prove that $W_t(q)$ is strictly increasing in q for $t \geq 1$. To see why, first assume that t is odd, and note that $W_t(q') - W_t(q)$ equals

$$(t-k)\binom{t}{k}\int_{1-q'}^{1-q} x^{t-k-1}(1-x)^k dx,$$

which is strictly positive if $q' > q$. Next, assume that $t \geq 2$ is even. By Lemma 4, $W_{t-1}(q) = W_t(q)$, and it follows that $W_t(q') - W_t(q) = W_{t-1}(q') - W_{t-1}(q) > 0$ for $q' > q$. To complete the proof, it is enough to show that

$$\sum_{t=0}^{N}\binom{N}{t}p^t(1-p)^{N-t}W_t(q') < \sum_{t=0}^{N}\binom{N}{t}(p')^t(1-p')^{N-t}W_t(q')$$

for $p < p'$. But this is true because an increase in p results in a new distribution of t that first order stochatically dominates the previous one, and by Lemma 4, $W_t(q')$ is monotonically increasing in t such that $W_t(q') \leq W_{t+1}(q')$.

Since our strategy construction strictly increases p and q in the notation of Lemmas 4 and 5 in at least one state of the world, we have a strict improvement over the non-cutoff strategy $\lambda_s^j(q)$ and Lemma 3 is proved.

We can now prove Proposition 1. First, note that by what McLennan (1998) has shown the optimal symmetric strategy, if it exists, must be an equilibrium strategy. In Lemma 3, we have shown the existence of the optimal strategy and that it cannot have a non-cutoff form. Next, consider the optimal equilibrium outcome in a biased electorate under unawareness regarding perception biases. First, assume that the outcome is not consistent with a cutoff outcome; i.e., a positive measure of individuals with identical q and s will behave in different ways due to perceptions biases. But this is inconsistent with the optimal strategy outcome, and thus worse than the optimal equilibrium outcome in an unbiased electorate. Second, assume that the outcome is consistent with a cutoff behavior outcome due to the particular form of overconfidence and underconfidence biases. If the outcome involves no interior cutoff, then this is surely suboptimal under the conditions stated in Lemma 2. If there is an interior cutoff involved, this is a nondegenerate case as not only the outcome should be consistent with another optimal equilibrium (to rule out the harm of perception biases) but also the form of $\lambda_o(q)$ must be very specific; for example, $\lambda_o(q) = 1$ must hold in a non-singleton interval, thereby ruling out underconfidence and unbiased perceptions in that interval as well as the possibility of different levels of overconfidence. Thus, a small perturbation in $\lambda_o(q)$ (or a small perturbation in π and $p_o(q)$[47]) is enough to rule out this last case. As a result, behavior with biased perceptions is generically inconsistent with the optimal equilibrium outcome in the unbiased case and thus suboptimal under the conditions stated in Lemma 2.

As mentioned in the main text all of our results and proofs are robust to assuming that there are several overconfidence functions $p_o^j(q) \in (q, \bar{q}]$ and respective probabilities $\lambda_o^j(q)$ where $p_o^j(q) < p_o^{j+1}(q)$ for every q and $j \in \{1, ..., J-1\}$.

Corollary 1: *Under conditions stated in Lemma 2, a responsive equilibrium always exists and has at least one interior cutoff.*

Proof: Given Lemma 3, an optimal symmetric equilibrium always exists, and under the conditions stated in Lemma 2, the optimal symmetric equilibrium involves an interior cutoff (i.e., it is a resposive equilibrium). Thus, there exists a responsive equilibrium under the conditions stated in Lemma 2.

A.2 Results in Section 2.1 under awareness of others' perception biases

Here, we assume that every i knows that for every j and q_j, $p_j(q_j)$ takes one of three possible values: $p_o(q_j)$, $p_u(q_j)$, and q_j with respective probabilities $\lambda_o(q)$, $\lambda_u(q)$, and $1-\lambda_o(q)-\lambda_u(q)$.

[47] For every π value except possibly those with a Lebesque measure of zero, there can exist only countably many equilibria (and only a subset will be optimal). Thus, perturbing π and $p_o(q)$ also suffice to rule out nongeneric cases.

Under this assumption, the first part of Lemma 1 naturally holds. Also, our proof for showing that $q^a = \Pr(S = B|piv_a)$ and $q^b = \Pr(S = A|piv_b)$ if the correct policy is chosen with a probability (weakly) greater than 0.5 in either state in equilibrium is unaffected. However, it may no longer be true that if $\pi = \underline{q} = 0.5$, then $q^a = q^b$ in every equilibrium or if $\pi = 0.5$, then $q^a = \Pr(S = B|piv_a)$. Lemmas 2 and 3 hold with awareness, and under the conditions stated in Lemma 2, the optimal equilibrium strategy with perception biases and awareness must involve at least one interior cutoff, since the steps used in the proof of Lemma 2 to show equilibria without interior cutoffs are suboptimal still apply. Thus, Proposition 1 holds, and all the main results are robust to awareness regarding others' overconfidence and underconfidence biases.

A.3 Proofs of Results in Section 2.3

We will invoke the following Lemma in our proofs.

Lemma 6: *In a symmetric equilibrium of every model discussed in the main text with independent signals, $q^a = \Pr(S = B|piv_a)$ and $q^b = \Pr(S = A|piv_b)$ must hold if the correct policy is chosen in either state with a probability (weakly) greater than* 0.5*.*

Proof: We proved the same statement in the proof of Lemma 1, and it can be seen that no argument depends on the distribution of signal precisions being identical or the absence of partisans. In particular, the difference in the distributions of signal precisions across states or the presence of partisans are accounted for in the probabilities of piv_a and piv_b events. Thus the proof presented in Lemma 1 applies to both extensions.

Proof of Proposition 2 First, we show that the sequence of optimal equilibria in the unbiased equilibrium is such that the correct policy is chosen with a probability that goes to one as $N \to \infty$. To show this, first note that the proof in Lemma 3 applies to show the existence of an optimal symmetric strategy and this strategy has a cutoff form. Then, it is enough to show that there exists a symmetric strategy which results in the correct policy being chosen with a probability that goes to 1 as $N \to \infty$ in either state (the optimal equilibrium exists by Lemma 3 and can never do worse). To show the existence of such a strategy, we assume without loss of generality that the (first) condition in the Proposition statement for overconfident electorates holds. Then, there exists a $q^* \in (\underline{q}, \bar{q})$ such that every non-partisan with $s_i = \alpha$ or with $s_i = \beta$ and $q_i < q^*$ votes for a, and every non-partisan with $s_i = \beta$ and $q_i > q^*$ votes for b such that the following holds: $p_a + (1 -$

$p_b)\left(\int_{\underline{q}}^{\bar{q}} qdF + \int_{\underline{q}}^{q^*}(1-q)dF\right) > p_b + (1-p_a-p_b)\int_{q^*}^{\bar{q}}(1-q)dF$, and $p_a + (1-p_a-p_b)\left(\int_{\underline{q}}^{\bar{q}}(1-q)dF + \int_{\underline{q}}^{q^*} qdF\right) < p_b + (1-p_a-p_b)\int_{q^*}^{\bar{q}} qdF$. The existence of this q^* follows from the fact that there exists a $\hat{q} \in (\underline{q}, \bar{q})$ such that $p_a + (1-p_a-p_b)\left(\int_{\underline{q}}^{\bar{q}} qdF + \int_{\underline{q}}^{\hat{q}}(1-q)dF\right) = p_b + (1-p_a-p_b)\int_{\hat{q}}^{\bar{q}}(1-q)dF$ and as a result, both inequalities are satisfied setting $q^* = \hat{q} + \epsilon$, where is $\epsilon > 0$ arbitrarily small. The equality above is due to continuity and the condition in the statement of the proposition: the left hand side of the equality is strictly greater (smaller) than the right hand side if ($\hat{q} = \bar{q}$) $\hat{q} = \underline{q}$. If $q^* = \hat{q}$, the latter inequality is automatically satisfied because $\int_{\underline{q}}^{\bar{q}} q > 0.5$ (as implied by the condition spelled out in the proposition). This strategy implies that the relative turnout share for the correct policy is strictly greater than 0.5 in either state ensuring that the correct policy is chosen with a probability that goes to one in either state.

Next, we characterize those equilibria in which the correct policy is chosen with a probability that goes to one as as $N \to \infty$ in both states. We initially assume away perception biases for ease of notation, but this is without loss of generality for the characterization under unawareness. Let q_N^a and q_N^b denote the respective equilibrium cutoffs for electorate size N. We will use the following claim.

Claim 1: *Assume that $\bar{q} \geq 1 - \underline{q}$ and consider those equilibria in which the correct policy is chosen with a probability that goes to one in both states. Then, $q_N^b < \bar{q}$ in large elections and* $\limsup_{N\to\infty} q_N^b < \bar{q}$.

Proof: First, consider the case in which $\bar{q} < 1$ and assume towards a contradiction that $\limsup_{N\to\infty} q_N^b \geq \bar{q}$. By Lemma 6, this implies that $\limsup_{N\to\infty} \frac{\pi \Pr(piv_b|A,N)}{(1-\pi)\Pr(piv_b|B,N)} \geq \frac{\bar{q}}{1-\bar{q}}$. Then, (with an abuse of notation) there exists an electorate size N and equilibrium sequence (q_N^a, q_N^b) such that $\frac{(1-\pi)\Pr(piv_b|B,N)}{\pi \Pr(piv_b|A,N)}$ goes to a number weakly smaller than $\frac{1-\bar{q}}{\bar{q}}$. Moreover, $\frac{(1-\pi)\Pr(piv_a|B,N)}{\pi \Pr(piv_a|A,N)} \to \frac{q^a}{1-q^a}$, where $q^a = \lim_{N\to\infty} q_N^a \geq 0$ (taking a convergent subsequence of the subsequence if necessary). Note that $q^a = \lim_{N\to\infty} q_N^a < \bar{q}$ as otherwise policy b would win in both states with a probability that goes to one as $N \to \infty$, which is suboptimal. $\Pr(piv_a|B,N)$ and $\Pr(piv_b|B,N)$ have a common term (representing the tie events) and the other term differs by only a multiplicative term. That is, $\Pr(piv_a|B,N)$ can be written as $\Pr(piv_a|B,N) = x_N + y_N$, and thus, $\Pr(piv_b|B,N)$ is equal to $x_N + y_N\gamma_N$ where

$$\gamma_N = \frac{p_a + (1-p_a-p_b)\left(\int_{q_N^a}^{\bar{q}}(1-q)dF + \int_{\underline{q}}^{1-q_N^a} qdF\right)}{p_b + (1-p_a-p_b)\left(\int_{q_N^b}^{\bar{q}} qdF + \int_{\underline{q}}^{1-q_N^b}(1-q)dF\right)}.$$

We assume without loss of generality that $1 - q_N^a > \underline{q}$ and $q_N^b < \bar{q}$ for large N (the proof is virtually unaffected if $1-q_N^a \le \underline{q}$ or $q_N^b \ge \bar{q}$). However, since $q_N^b \to \bar{q}$ and $\bar{q} > 1-\underline{q}$, $1-q_N^b < \underline{q}$ must hold for all large N. $\Pr(piv_a|A,N)$ and $\Pr(piv_b|A,N)$ also have a common term, and the other term differs by a multiplicative term. Thus, $\Pr(piv_a|A,N)$ can be written as $\Pr(piv_a|A,N) = w_N + z_N$, and $\Pr(piv_b|A,N)$ is equal to $w_N + z_N\xi_N$ where

$$\xi_N = \frac{p_a + (1-p_a-p_b)\left(\int_{q_N^a}^{\bar{q}} qdF + \int_{\underline{q}}^{1-q_N^a}(1-q)dF\right)}{p_b + (1-p_a-p_b)\left(\int_{q_N^b}^{\bar{q}}(1-q)dF + \int_{\underline{q}}^{1-q_N^b} qdF\right)}.$$

As the correct policy must be chosen with a probability that goes to one in both states, $\xi_N > 1$ and $\gamma_N < 1$ must hold. Moreover, if $\bar{q} < 1$, then $q_N^a > 0.5$ must hold because otherwise $q_N^b \to \bar{q}$ and $p_a + (1-p_a-p_b)\int_{0.5}^{\bar{q}}(1-q)dF > p_b$ imply that policy a is chosen in both states with a probability that goes to one. Thus, $q^a = \lim_{N\to\infty} q_N^a \in (0.5, \bar{q})$. As a result,

$$\begin{aligned} \frac{\bar{q}}{1-\bar{q}}\frac{q^a}{1-q^a} &= \lim_{N\to\infty} \frac{\frac{1-\pi}{\pi}\frac{x_N+y_N}{w_N+z_N}}{\frac{1-\pi}{\pi}\frac{x_N+\gamma_N y_N}{w_N+\xi_N z_N}} = \lim_{N\to\infty} \frac{x_N+y_N}{x_N+\gamma_N y_N}\frac{w_N+\xi_N z_N}{w_N+z_N} \\ &\le \limsup_{N\to\infty} \frac{\xi_N}{\gamma_N} = \frac{p_a + (1-p_a-p_b)\left(\int_{q^a}^{\bar{q}} qdF + \int_{\underline{q}}^{1-q^a}(1-q)dF\right)}{p_a + (1-p_a-p_b)\left(\int_{q^a}^{\bar{q}}(1-q)dF + \int_{\underline{q}}^{1-q^a} qdF\right)}. \end{aligned}$$

However, the very last term is strictly smaller than $\frac{\int_{q^a}^{\bar{q}} qdF + \int_{\underline{q}}^{1-q^a}(1-q)dF}{\int_{q^a}^{\bar{q}}(1-q)dF + \int_{\underline{q}}^{1-q^a} qdF}$, which in turn is strictly smaller than $\frac{\bar{q}}{1-\bar{q}}$, resulting in a contradiction. Thus, $\limsup_{N\to\infty} q_N^b < \bar{q} < 1$. We now analyze the case in which $\bar{q} = 1$. Assume towards a contradiction that $\limsup_{N\to\infty} q_N^b \ge 1$. This implies that $\limsup_{N\to\infty} \frac{\pi \Pr(piv_b|A,N)}{(1-\pi)\Pr(piv_b|B,N)} = \infty$. Abusing notation, we have that $\frac{\Pr(piv_b|B,N)}{\Pr(piv_b|A,N)} \to 0$. However, if $\frac{\Pr(piv_b|B,N)}{\Pr(piv_b|A,N)} \to 0$, then $\frac{\Pr(piv_a|B,N)}{\Pr(piv_a|A,N)} \to 0$ must also hold. This is because $\liminf_{N\to\infty}\gamma_N > 0$ and $\limsup_{N\to\infty}\xi_N < \infty$, and $\frac{\Pr(piv_b|B,N)}{\Pr(piv_b|A,N)} = \frac{x_N+\gamma_N y_N}{w_N+\xi_N z_N} \ge \frac{\gamma_N}{\xi_N}\frac{x_N+y_N}{w_N+z_N} = \frac{\gamma_N}{\xi_N}\frac{\Pr(piv_a|B,N)}{\Pr(piv_a|A,N)}$. In turn, $q_N^a \to 0$. As a result, almost everyone but the b-partisans votes for policy a in large elections, and a is chosen in both states with a probability that goes to one, a contradiction.

Without awareness: We now show that a high enough degree of overconfidence is harmful in this setting. In large elections, a sufficiently high level of Dunning-Kruger effect implies that

$$p_a+(1-p_a-p_b)\left(\int_{q^a}^{\bar{q}} qdF + \int_{p_o^{-1}(q^a)}^{q^a}\lambda_o(q)qdF + \int_{\underline{q}}^{p_o^{-1}(1-q^a)}(1-q)dF + \int_{p_o^{-1}(1-q^a)}^{1-q^a}(1-\lambda_o(q))(1-q)dF\right)$$

is strictly lower than

$$p_b+(1-p_a-p_b)\left(\int_{q^b}^{\bar{q}}(1-q)dF + \int_{p_o^{-1}(q^b)}^{q^b}\lambda_o(q)(1-q)dF + \int_{p_o^{-1}(1-q^b)}^{1-q^b} qdF + \int_{p_o^{-1}(1-q^b)}^{1-q^b}\lambda_o(q)qdF\right).$$

To see why, note that if in the optimal equilibrium $1 - q^a > \underline{q}$ and/or $1 - q^b > \underline{q}$, then a high level of Dunning Kruger effect makes these cutoffs trivial. Moreover, a sufficiently high level of Dunning Kruger effect makes $p_o^{-1}(q^b)$ closer and closer to $\underline{q}$ substantially increasing turnout from low skill people in accordance with their (low quality) news signal, and as a result, the condition in the statement of the Proposition ensures that policy b is chosen in both states with a probability that goes to one as $N \to \infty$ (e.g., if $p_o(\underline{q})$ is close to or greater than $\limsup_{N\to\infty} q_N^b$, and $\lambda_o(q)$ is sufficiently high at every $q \in (\underline{q}, \limsup_{N\to\infty} q_N^b)$, then policy b is chosen in both states with a probability that goes to one).

With awareness: We consider a finite type space with $\{q_1, q_2, ..., q_T\}$ where $q_T = 1$. Obviously $q_N^b < q_T = 1$ for every N in any responsive equilibrium. As a result, a sufficiently high and widespread level of Dunning Kruger effect in the population will prevent information aggregation despite awareness: if for example $p_o(q_1) = q_T$, then the sufficient condition ensures that there exist high enough $\{\lambda_o(q_t)\}_{t\in\{1,2,...,T-1\}}$ such that

$$\frac{p_b - p_a}{1 - p_a - p_b} \geq \sum_{t=1}^{T} q_t + (1 - \lambda_o(q_t))(1 - q_t)\Pr(q_t) - \sum_{t=1}^{T} \lambda_o(q_t)(1 - q_t)\Pr(q_t),$$

and thus, expected turnout for b exceeds expected turnout for a in state A (i.e., even if we assume every unbiased i with $q_i < q_T$ votes against b regardless of signal).

Proof of Proposition 3 Let q^a and q^b be as defined in (2) and (3) respectively. In every Bayesian Nash equilibrium, individual i votes for a if either i's signal is α and $\frac{q_i}{q_i+1-m(q_i)} \geq q^a$ or i's signal is β and $\frac{m(q_i)}{m(q_i)+1-q_i} \leq 1 - q^a$. In a similar vein, i votes for b if either i's signal is β and $\frac{m(q_i)}{m(q_i)+1-q_i} \geq q^b$ or i's signal is α and $\frac{q_i}{q_i+1-m(q_i)} \leq 1 - q^b$. One issue is the existence of an optimal strategy. We will either assume that one exists or assume that there is a finite set of types, in which case an optimal symmetric strategy always exists and must coincide with the optimal symmetric equilibrium. The equilibrium characterization and arguments below are virtually unaffected if there is a finite set of q types.[48] We now construct a strategy that fully aggregates information in the limit in the absence of perception biases. The strategy is such that i votes for the policy that matches s_i if $q_i \geq q^*$ and abstains otherwise. If we select cutoff q^* such that $\int_{q^*}^{\bar{q}} m(q)dF > \int_{q^*}^{\bar{q}} (1 - m(q))dF$ (such q^* surely exists as $m(\bar{q}) > 0.5$) and

[48] To be more precise, if there is a type q (or rather $p(q)$) that exactly equals one of the equilibrium cutoffs, that type may be randomizing in equilibrium. For example, if i's signal is α and it turns out that $q_i = q^a$, then i may randomize in equilibrium. However, such randomization will be accounted for in the pivotality calculus, and the formal equilibrium characterization is unaffected.

$\int_{q^*}^{\bar{q}} q dF > \int_{q^*}^{\bar{q}} (1-q) dF$ hold, then this implies that the relative turnout share for the correct policy is strictly greater than 0.5 in either state ensuring that the correct policy is chosen in both states with a probability that goes to one as $N \to \infty$. Thus, the optimal equilibrium will result in the same in the limit.

Next, we characterize those equilibria in which the correct policy is chosen with a probability that goes to one as as $N \to \infty$ in both states. We initially assume away perception biases for ease of notation, but this is without loss of generality for the characterization under unawareness of perception biases. The case with awareness is considered at the end of the proof. Let q_N^a and q_N^b denote the respective equilibrium cutoffs for electorate size N. We first assume that $\bar{q} < 1$. By the assumption that $m(\bar{q}) > 1 - m(\underline{q})$, we have that $\frac{m(\bar{q})}{m(\bar{q})+1-\bar{q}} > \frac{1-m(\underline{q})}{\underline{q}+1-m(\underline{q})}$ and that $\bar{q} > 1 - \underline{q}$. Furthermore, $\bar{q} > 1 - \underline{q}$ and $m(\bar{q}) > 1 - m(\underline{q})$ imply that $\frac{\bar{q}}{\bar{q}+1-m(\bar{q})} > \frac{1-\underline{q}}{m(\underline{q})+1-\underline{q}}$. We now show that $\limsup_{N\to\infty} q_N^a < \frac{\bar{q}}{\bar{q}+1-m(\bar{q})}$ and $\limsup_{N\to\infty} q_N^b < \frac{m(\bar{q})}{m(\bar{q})+1-\bar{q}}$ must hold. Assume towards a contradiction that $\limsup_{N\to\infty} q_N^b \geq \frac{m(\bar{q})}{m(\bar{q})+1-\bar{q}}$. Then, by the condition $\frac{m(\bar{q})}{m(\bar{q})+1-\bar{q}} > \frac{1-m(\underline{q})}{\underline{q}+1-m(\underline{q})}$, turnout rate for policy b goes to zero as $N \to \infty$. This implies that $\limsup_{N\to\infty} q_N^a \geq \frac{\bar{q}}{\bar{q}+1-m(\bar{q})}$ must hold because, otherwise, relative turnout rate for policy b (hence, the probability that b is chosen) goes to zero in state B, which contradicts our initial hypothesis. More generally, $\limsup_{N\to\infty} q_N^b \geq \frac{m(\bar{q})}{m(\bar{q})+1-\bar{q}}$ if and only if $\limsup_{N\to\infty} q_N^a \geq \frac{\bar{q}}{\bar{q}+1-m(\bar{q})}$. Strict inequalities can however not hold, as that would mean at least one policy receives zero votes in both states as $N \to \infty$, which contradicts our initial hypothesis that the correct policy is chosen in both states with a probability that goes to one. Hence, $\limsup_{N\to\infty} \frac{(1-\pi)\Pr(piv_a|B,N)}{\pi \Pr(piv_a|A,N)} = \frac{\bar{q}}{1-m(\bar{q})}$ and $\limsup_{N\to\infty} \frac{(1-\pi)\Pr(piv_b|B,N)}{\pi \Pr(piv_b|A,N)} = \frac{1-\bar{q}}{m(\bar{q})}$. Thus, (with an abuse of notation) there exists an electorate size N and equilibrium sequence $\{q_N^a, q_N^b\}$ such that $\frac{(1-\pi)\Pr(piv_a|B,N)}{\pi \Pr(piv_a|A,N)} \to \frac{\bar{q}}{1-m(\bar{q})}$ and $\frac{(1-\pi)\Pr(piv_b|B,N)}{\pi \Pr(piv_b|A,N)} \to \frac{1-\bar{q}}{m(\bar{q})}$. We will now show that these limits result in a contradiction. To see why, note that $\Pr(piv_a|B,N)$ and $\Pr(piv_b|B,N)$ have a common term (representing the tie events) and the other term differs by only a multiplicative term. That is, $\Pr(piv_a|B,N)$ can be written as $\Pr(piv_a|B,N) = x_N + y_N$, and thus, $\Pr(piv_b|B,N)$ is equal to $x_N + y_N\gamma_N$ where $\gamma_N = \frac{\int_{\hat{q}_N^a}^{\bar{q}} (1-m(q))dF}{\int_{\hat{q}_N^b}^{\bar{q}} m(q)dF}$, and $\hat{q}_N^a$ and $\hat{q}_N^b$ the respective values that solve for $q_N^a = \frac{\hat{q}_N^a}{\hat{q}_N^a+1-m(\hat{q}_N^a)}$ and $q_N^b = \frac{m(\hat{q}_N^b)}{m(\hat{q}_N^b)+1-\hat{q}_N^b}$.[49] In a similar vein, $\Pr(piv_a|A,N)$ and $\Pr(piv_b|A,N)$ have a common term, and the other term differs by a

[49] By the inequalities we have shown above, if $\hat{q}_N^a$ and $\hat{q}_N^b$ are very close to $\frac{\bar{q}}{\bar{q}+1-m(\bar{q})}$ and $\frac{m(\bar{q})}{m(\bar{q})+1-\bar{q}}$ respectively, then $1-\hat{q}_N^a$ and $1-\hat{q}_N^b$ cutoffs are irrelevant and voting against signal will not take place because $1-\hat{q}_N^a < \frac{m(\underline{q})}{m(\underline{q})+1-\underline{q}}$.and $1-\hat{q}_N^b < \frac{\underline{q}}{\underline{q}+1-m(\underline{q})}$ will hold.

multiplicative term. Thus, $\Pr(piv_a|A,N)$ can be written as $\Pr(piv_a|A,N) = w_N + z_N$, and $\Pr(piv_b|A,N)$ is equal to $w_N + z_N\xi_N$ where $\xi_N = \frac{\int_{\hat{q}_N^a}^{\bar{q}} qdF}{\int_{\hat{q}_N^b}^{\bar{q}} (1-q)dF}$.

Note that $\xi_N > 1$ and $\gamma_N < 1$ must hold as we focus on those equilibria in which the correct policy is chosen in both states with a probability that goes to one as $N \to \infty$. Moreover, the term ξ_N is bounded above because $\xi_N \leq \frac{\bar{q}}{1-\bar{q}} \frac{1-F(\hat{q}_N^a)}{1-F(\hat{q}_N^b)}$ and $1 > \gamma_N \geq \frac{1-m(\bar{q})}{m(\bar{q})} \frac{1-F(\hat{q}_N^a)}{1-F(\hat{q}_N^b)}$ imply that $\xi_N \leq \frac{\bar{q}}{1-\bar{q}} \frac{m(\bar{q})}{1-m(\bar{q})} < \infty$ as $\bar{q} < 1$. Therefore, $\limsup_{N\to\infty} \xi_N = \xi^* = \frac{\bar{q}}{1-\bar{q}} \limsup_{N\to\infty} \frac{1-F(\hat{q}_N^a)}{1-F(\hat{q}_N^b)} < \infty$. Moreover, $\liminf_{N\to\infty} \gamma_N = \gamma^* = \frac{1-m(\bar{q})}{m(\bar{q})} \liminf_{N\to\infty} \frac{1-F(\hat{q}_N^a)}{1-F(\hat{q}_N^b)} > 0$ because $\frac{1-F(\hat{q}_N^a)}{1-F(\hat{q}_N^b)} > \frac{1-\bar{q}}{\bar{q}} > 0$ from $\xi_N > 1$ and $\bar{q} < 1$. Since $\frac{(1-\pi)\Pr(piv_a|B,N)}{\pi \Pr(piv_a|A,N)} \to \frac{\bar{q}}{1-m(\bar{q})}$ by hypothesis, $\lim_{N\to\infty} \frac{(1-\pi)(x_N+y_N)}{\pi(w_N+z_N)} = \frac{\bar{q}}{1-m(\bar{q})}$. We need the following claim.

Claim 2: *Assume that $\bar{q} < 1$ and consider a sequence of N such that $\xi_N > 1 > \gamma_N$ and equilibrium abstention rate is bounded away from 0 (i.e., it is not necessary that $\hat{q}_N^a \to \bar{q}$ and $\hat{q}_N^b \to \bar{q}$). Under these conditions* $\liminf_{N\to\infty} x_N/y_N > 0$ *and* $\limsup_{N\to\infty} x_N/y_N < \infty$. *Similarly,* $\liminf_{N\to\infty} w_N/z_N > 0$ *and* $\limsup_{N\to\infty} w_N/z_N < \infty$ *must hold.*

Proof: It is enough to prove that $\liminf_{N\to\infty} x_N/y_N > 0$ and $\limsup_{N\to\infty} x_N/y_N < \infty$ as the proof of the other statement is analogous. First, we prove that $\liminf_{N\to\infty} x_N/y_N > 0$. Suppose towards a contradiction that there exists a sequence which we again denote by N such that x_N/y_N goes to 0. Thus, $x_N/\gamma_N y_N$ goes to 0 as well because $\gamma^* > 0$ as we showed above. Let $\Pr(t_a, t_b|B,N)$ denote the probability that there are t_a votes for policy a and t_b votes for policy b in state B with electorate size N. It can be checked that $\Pr(t+1,t|B,N) = \Pr(t,t|B,N) \frac{N-2t}{t+1} \frac{T_a(B,N)}{1-(T_a(B,N)+T_b(B,N))}$, where $T_j(S,N)$ denotes the expected turnout rate for policy $j \in \{a,b\}$ in state $S \in \{A,B\}$ with electorate size N (thus, $\gamma_N = \frac{T_a(B,N)}{T_b(B,N)}$). If $NT_b(B,N)$ goes to a finite number as $N \to \infty$, then obviously $x_N/\gamma_N y_N$ cannot go to 0 because it is greater than $\frac{1-(T_a(B,N)+T_b(B,N))}{NT_a(B,N)}$, which is bounded below by a number strictly larger than zero for any N because by hypothesis $\gamma_N < 1$ and thus, $T_a(B,N) < T_b(B,N)$. Next, assume that $NT_b(B,N)$ goes to infinity as $N \to \infty$. Note that $\Pr(t,t|B) = \Pr(t,t-1|B) \frac{N-2t+1}{t} \frac{T_b(B,N)}{1-(T_a(B,N)+T_b(B,N))}$, and that since $NT_b(B,N)$ goes to infinity as $N \to \infty$,there must exist $t^*(N) > 0$ such that $\frac{N-2t^*(N)+1}{t^*(N)} \frac{T_b(B,N)}{1-(T_a(B,N)+T_b(B,N))} > 1 \geq \frac{N-2t^*(N)-1}{t^*(N)+1} \frac{T_b(B,N)}{1-(T_a(B,N)+T_b(B,N))}$. In particular, for every $t \geq t^*(N)$, $\frac{N-2t}{t+1} \frac{T_a(B,N)}{1-(T_a(B,N)+T_b(B,N))}$ is bounded above by $1 + \frac{1}{t^*(N)+1} \frac{T_a(B,N)}{1-(T_a(B,N)+T_b(B,N))}$, which is a finite number because abstention rate is bounded away from 0 by hypothesis. Recalling that $\Pr(t+1,t|B,N) = \Pr(t,t|B,N) \frac{N-2t}{t+1} \frac{T_a(B,N)}{1-(T_a(B,N)+T_b(B,N))}$, x_N/y_N cannot go to 0, a contradiction.

We now show that $\limsup_{N\to\infty} x_N/y_N < \infty$. Suppose towards a contradiction that there exists a sequence which we again denote by N such that y_N/x_N goes to 0. Note that $\Pr(t,t+1|B,N) = \Pr(t,t|B,N)\frac{N-2t}{t+1}\frac{T_b(B,N)}{1-(T_a(B,N)+T_b(B,N))}$, and $\Pr(t+1,t+1|B,N) = \Pr(t,t+1|B,N)\frac{N-2t-1}{t+1}\frac{T_a(B,N)}{1-(T_a(B,N)+T_b(B,N))}$. If $NT_b(B,N)$ goes to a finite number as $N \to \infty$, then y_N/x_N cannot go to 0 because $\frac{y_N}{x_N} > \frac{1}{2}\min\{\frac{1-(T_a(B,N)+T_b(B,N))}{(N-1)T_a(B,N)}, \frac{NT_b(B,N)}{1-(T_a(B,N)+T_b(B,N))}\}$, which converges to a number strictly larger than zero ($NT_b(B,N)$ cannot go to 0 as $N \to \infty$). Next, assume that $NT_b(B,N)$ goes to infinity as $N \to \infty$. Then, for large N there must exist $t^*(N) \geq 1$ such that $\frac{N-2t^*(N)}{t^*(N)+1}\frac{T_b(B,N)}{1-(T_a(B,N)+T_b(B,N))} > 1$ and $\frac{N-2t}{t+1}\frac{T_b(B,N)}{1-(T_a(B,N)+T_b(B,N))} \leq 1$ for $t > t^*(N)$. Thus, $\frac{N-2t-1}{t+1}\frac{T_a(B,N)}{1-(T_a(B,N)+T_b(B,N))} < 1$ for $t > t^*(N)$. This proves that y_N/x_N cannot go to 0, a contradiction. Hence, Claim 2 is proved.

We now prove that $\frac{(1-\pi)\Pr(piv_a|B,N)}{\pi\Pr(piv_a|A,N)} \to \frac{\bar{q}}{1-m(\bar{q})}$ and $\frac{(1-\pi)\Pr(piv_b|B,N)}{\pi\Pr(piv_b|A,N)} \to \frac{1-\bar{q}}{m(\bar{q})}$ result in a contradiction. Taking a convergent subsequence of x_N/y_N first and then a convergent subsequence of w_N/z_N if necessary, we can write

$$\lim_{N\to\infty}\frac{(1-\pi)(x_N+y_N)}{\pi(w_N+z_N)} = \lim_{N\to\infty}\frac{(1-\pi)(x_N/y_N+1)}{\pi(w_N/z_N+1)}\lim_{N\to\infty}\frac{y_N}{z_N} = \frac{(1-\pi)(c+1)}{\pi(d+1)}\lim_{N\to\infty}\frac{y_N}{z_N},$$

where, abusing notation, $c = \lim_{N\to\infty}\frac{x_N}{y_N} \in (0,\infty)$ and $d = \lim_{N\to\infty}\frac{w_N}{z_N} \in (0,\infty)$ by Claim 2. Thus, $\lim_{N\to\infty} y_N/z_N$ exists and $\lim_{N\to\infty} y_N/z_N \in (0,\infty)$ since $\lim_{N\to\infty}\frac{(1-\pi)(x_N/y_N+1)}{\pi(w_N/z_N+1)} = \frac{(1-\pi)(c+1)}{\pi(d+1)} \in (0,\infty)$ and $\lim_{N\to\infty}\frac{(1-\pi)(x_N+y_N)}{\pi(w_N+z_N)} = \frac{\bar{q}}{1-m(\bar{q})} \in (0,\infty)$. Thus, we have that

$$\frac{\bar{q}}{1-m(\bar{q})} = \frac{(1-\pi)(c+1)}{\pi(d+1)}\lim_{N\to\infty}\frac{y_N}{z_N},$$

and from $\lim_{N\to\infty}\frac{(1-\pi)(x_N+\gamma_N y_N)}{\pi(w_N+\xi_N z_N)} = \frac{1-\bar{q}}{m(\bar{q})}$, we have that

$$\frac{1-\bar{q}}{m(\bar{q})} = \frac{(1-\pi)(c+\lim_{N\to\infty}\gamma_N)}{\pi(d+\lim_{N\to\infty}\xi_N)}\lim_{N\to\infty}\frac{y_N}{z_N}.$$

taking convergent subsequences of γ_N and ξ_N subsequences if necessary. Note that $\lim_{N\to\infty}\frac{\xi_N}{\gamma_N} = \frac{\bar{q}}{1-\bar{q}}\frac{m(\bar{q})}{1-m(\bar{q})}$ and $\lim_{N\to\infty}\gamma_N \leq 1 \leq \lim_{N\to\infty}\xi_N$ (with at least one inequality being strict). Thus, the two equalities above cannot be satisfied at the same time. As a result, there must exist a $\hat{q} < \bar{q}$ such that $\hat{q}_N^a$ and $\hat{q}_N^b$ are smaller than $\hat{q}$ for all N.

For the case in which $\bar{q} = 1$, we assume that every i is an a-partisan with probability $\frac{p}{2}$ and a b-partisan with probability $\frac{p}{2}$ where $p > 0$ is possibly small. (The assumption that every i is an a-partisan with probability $\frac{p}{2}$ and a b-partisan with probability $\frac{p}{2}$ also works with $\bar{q} < 1$ as in this case $\limsup_{N\to\infty}\frac{\xi_N}{\gamma_N} < \frac{\bar{q}}{1-\bar{q}}\frac{m(\bar{q})}{1-m(\bar{q})}$.) Assume towards a contradiction that $\limsup_{N\to\infty} q_N^b \geq \frac{m(\bar{q})}{m(\bar{q})+1-\bar{q}} = 1$. As in the case with $\bar{q} < 1$, $\limsup_{N\to\infty} q_N^b \geq 1$ if and only if $\limsup_{N\to\infty} q_N^a \geq \frac{\bar{q}}{\bar{q}+1-m(\bar{q})}$. In particular, $\limsup_{N\to\infty} q_N^a = \frac{\bar{q}}{\bar{q}+1-m(\bar{q})}$ and $\limsup_{N\to\infty} q_N^b = 1$

as strict inequalities cannot hold. These imply that $\limsup_{N\to\infty} \frac{(1-\pi)\Pr(piv_a|B,N)}{\pi\Pr(piv_a|A,N)} = \frac{1}{1-m(\bar{q})}$ and $\limsup_{N\to\infty} \frac{(1-\pi)\Pr(piv_b|B,N)}{\pi\Pr(piv_b|A,N)} = 0$. Thus, (with an abuse of notation) there exists an equilibrium sequence q_N^a and q_N^b giving rise to $\frac{(1-\pi)\Pr(piv_a|B,N)}{\pi\Pr(piv_a|A,N)} \to \frac{1}{1-m(\bar{q})}$ and $\frac{(1-\pi)\Pr(piv_b|B,N)}{\pi\Pr(piv_b|A,N)} \to 0$ with $\hat{q}_N^a \to 1$ and $\hat{q}_N^b \to 1$. Using the previous notation introduced above, these limits translate to $\lim_{N\to\infty} \frac{(1-\pi)(x_N+y_N)}{\pi(w_N+z_N)} = \frac{1}{1-m(\bar{q})}$ and $\lim_{N\to\infty} \frac{(1-\pi)(x_N+\gamma_N y_N)}{\pi(w_N+\xi_N z_N)} = 0$. Thus, $\lim_{N\to\infty} \frac{x_N+y_N}{x_N+\gamma_N y_N}\frac{w_N+\xi_N z_N}{w_N+z_N} = \infty$. However, this is impossible because $\lim_{N\to\infty} \frac{x_N+y_N}{x_N+\gamma_N y_N}\frac{w_N+\xi_N z_N}{w_N+z_N} \leq \lim_{N\to\infty} \frac{\xi_N}{\gamma_N} = 1$ as $\hat{q}_N^a \to 1$ and $\hat{q}_N^b \to 1$,

$$\gamma_N = \frac{\frac{p}{2} + (1-p)\int_{\hat{q}_N^a}^1 (1-m(q))dF}{\frac{p}{2} + (1-p)\int_{\hat{q}_N^b}^1 m(q)dF},$$

and

$$\xi_N = \frac{\frac{p}{2} + (1-p)\int_{\hat{q}_N^a}^1 qdF}{\frac{p}{2} + (1-p)\int_{\hat{q}_N^b}^1 (1-q)dF}.$$

Without awareness: We have shown above that $\hat{q}_N^a$ is bounded above away from $\bar{q}$ for $\bar{q} \leq 1$ for any equilibrium sequence such that the correct policy is chosen with a probability that goes to one as as $N \to \infty$ in both states. Given the bound on the equilibrium cutoff $\hat{q}_N^a$, we can construct an overconfidence function (as we did in the proof of Proposition 2) such that the probability that policy b is chosen in state B goes to zero because too many people vote, and too many people vote for their signal (i.e. policy a) due to the fact that $\int_{\underline{q}}^{\bar{q}} m(q)dF < 0.5$.

With awareness: We consider a finite type space with $\{q_1, q_2, ..., q_T\}$ where $q_T = 1 = m(q_T)$. Obviously $q_N^b < q_T = 1$ for every N in any responsive equilibrium. As a result, a sufficiently high and widespread level of Dunning Kruger effect in the population will prevent information aggregation despite awareness: if for example $p_o(q_1) = q_T$, then the sufficient condition ensures that there exist high enough $\{\lambda_o(q_t)\}_{t\in\{1,2,...,T-1\}}$ such that

$$\sum_{t=1}^{T} m(q_t) + (1-\lambda_o(q_t))(1-m(q_t))\Pr(q_t) < \sum_{t=1}^{T} \lambda_o(q_t)(1-m(q_t))\Pr(q_t),$$

and thus, expected turnout for a exceeds expected turnout for b in state B (even if we assume every unbiased i with $q_i < q_T$ votes against a regardless of s_i).

Proof of Proposition 4 We will first show that every responsive equilibrium consists of cutoffs $\underline{q}^a$, $\bar{q}^a$, $\underline{q}^b$ and $\bar{q}^b$ and derive Lemma 8 for the general case below. We start by assuming that the signal of individual i is α. In that case, i prefers voting for a over abstention if and only if

$$\frac{1}{2}\left(\Pr(piv_a \cap S = A|s_i = \alpha) - \Pr(piv_a \cap S = B|s_i = \alpha)\right) \geq 0$$

as in the independent signal case. Different from the independent signal case, we must differentiate between the case where s_i comes from a high quality source and the case where it comes from a low quality source. Let $Q_i = H$ ($Q_i = L$) denote the case where s_i comes from a high quality (low quality) source. It can be checked that $\Pr(piv_a \cap S = A | s_i = \alpha)$ equals

$$\Pr(piv_a \cap S = A | \, (s_i = \alpha \cap Q_i = H) \cup (s_i = \alpha \cap Q_i = L)),$$

and thus, $\Pr(piv_a \cap S = A | s_i = \alpha)$ equals

$$\frac{\Pr(piv_a \cap S = A \cap s_i = \alpha \cap Q_i = H) + \Pr(piv_a \cap S = A \cap s_i = \alpha \cap Q_i = L)}{\Pr(s_i = \alpha)},$$

where $\Pr(piv_a \cap S = A \cap s_i = \alpha \cap Q_i = H)$ equals

$$q_i \pi \Pr(piv_a \cap s_i = \alpha | S = A \cap Q_i = H).$$

Thus, we have that i weakly prefers voting for a over abstention with $s_i = \alpha$ if and only if

$$p_i(q_i) \geq \frac{x}{x+y},$$

where

$$\begin{aligned} x &= (1-\pi)\Pr(piv_a \cap s_i = \alpha | S = B \cap Q_i = L) - \pi \Pr(piv_a \cap s_i = \alpha | S = A \cap Q_i = L) \\ y &= \pi \Pr(piv_a \cap s_i = \alpha | S = A \cap Q_i = H) - (1-\pi)\Pr(piv_a \cap s_i = \alpha | S = B \cap Q_i = H). \end{aligned}$$

Next, consider the (weak) preference of i for voting for a over voting for b after observing an α signal. This will be the case if and only if given $s_i = \alpha$ and (perceived) signal precision $p_i(q_i)$, $p_i(q_i) \geq \frac{x+w}{x+w+y+z}$, where x and y are as defined above and

$$\begin{aligned} w &= (1-\pi)\Pr(piv_b \cap s_i = \alpha | S = B \cap Q_i = L) - \pi \Pr(piv_b \cap s_i = \alpha | S = A \cap Q_i = L) \\ z &= \pi \Pr(piv_b \cap s_i = \alpha | S = A \cap Q_i = H) - (1-\pi)\Pr(piv_b \cap s_i = \alpha | S = B \cap Q_i = H). \end{aligned}$$

Hence, we derive the cutoff $\bar{q}^a$ for voting for policy a conditional on $s_i = \alpha$: $\bar{q}^a = \max\{\frac{x}{x+y}, \frac{x+w}{x+w+y+z}\}$. Next, we derive the cutoff $\underline{q}^b$ for voting for b conditional on $s_i = \alpha$. Note that i prefers voting for b over abstention and voting for a if and only if $p_i(q_i) \leq \min\{\frac{w}{w+z}, \frac{x+w}{x+w+y+z}\}$. Hence, we derive the cutoff $\underline{q}^b$ for voting for b conditional on $s_i = \alpha$: $\underline{q}^b = \min\{\frac{w}{w+z}, \frac{x+w}{x+w+y+z}\}$. The derivation of $\underline{q}^a$ and $\bar{q}^b$ are analogous and therefore omitted. Thus, we obtain Lemma 8.

Lemma 8 *In the correlated signal model, every responsive and symmetric Bayesian Nash equilibrium consists of four cutoffs $\underline{q}^a$, $\bar{q}^a$, $\underline{q}^b$ and $\bar{q}^b$ such that* (1) *an individual votes for a* *if and only if either i's signal is* α *and* $p_i(q_i) \geq \bar{q}^a$ *(provided that* $\bar{q}^a < \bar{q}$*) or i's signal is* β *and* $p_i(q_i) \leq \underline{q}^a$ *(provided that* $\underline{q}^a > \underline{q}$*) and* (2) *an individual votes for* b *if and only if either i's signal is* β *and* $p_i(q_i) \geq \bar{q}^b$ *(provided that* $\bar{q}^b < \bar{q}$*) or i's signal is* α *and* $p_i(q_i) \leq \underline{q}^b$ *(provided that* $\underline{q}^b > \underline{q}$*).*

In the optimal equilibrium with an unbiased electorate, (i) the probability of selecting the correct policy must go to one in both states as N goes to infinity if $q_H = 1$, and (ii) the probability goes to a number weakly larger than $\sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k} q_H^k (1-q_H)^{n-k}$ as N goes to infinity if $q_H < 1$. To show why this is the case, we will either assume that an optimal equilibrium exists or assume that there is a finite set of types, in which case an optimal symmetric strategy always exists and must coincide with the optimal symmetric equilibrium. The equilibrium characterization and arguments below are virtually unaffected if there is a finite set of q types.[50]

We now construct a simple symmetric strategy for the case with $q_H = 1$. It is enough to note that even though $\int_{\underline{q}}^{\bar{q}} q dF < \int_{\underline{q}}^{\bar{q}} (1-q) dF$ holds, there always exists $\hat{q} \in (\underline{q}, \bar{q})$ such that $\int_{\hat{q}}^{\bar{q}} q dF > \int_{\hat{q}}^{\bar{q}} (1-q) dF$ since $\bar{q} > 0.5$. This strategy ensures that the relative turnout share for the correct policy is strictly greater than 0.5 in either state ensuring that the correct policy is chosen in both states with a probability that goes to one as $N \to \infty$. Thus, the optimal equilibrium will result in the same in the limit. Next, we construct a symmetric strategy for the case in which $q_H < 1$ and $\bar{q} > \frac{n}{n+1}$. Even though $\frac{1}{n}\int_{\underline{q}}^{\bar{q}} q dF < \int_{\underline{q}}^{\bar{q}} (1-q) dF$ holds, there always exists $\hat{q} \in (\underline{q}, \bar{q})$ such that $\frac{1}{n}\int_{\hat{q}}^{\bar{q}} q dF > \int_{\hat{q}}^{\bar{q}} (1-q) dF$ since $\bar{q} > \frac{n}{n+1}$. This condition ensures that it is the majority among the signals of high quality news outlets that determine the voting outcome; for example, whenever a majority of high quality sources provide an α signal, then the relative turnout rate for policy a is strictly greater than 0.5. As a result, the probability of selecting the correct policy goes to $\sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k} q_H^k (1-q_H)^{n-k}$ in both states as N goes to infinity, and thus the probability in the optimal equilibrium is weakly larger than $\sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k} q_H^k (1-q_H)^{n-k}$ in the limit.

We will now characterize equilibria in large elections with $q_H < 1$ such that the probability of selecting the correct policy goes to a number weakly larger than $\sum_{k=\frac{n+1}{2}}^{n} \binom{n}{k} q_H^k (1-q_H)^{n-k}$ as N goes to infinity. Let x denote the realized number of high quality news outlets (out of a total of n) such that $s_h = \alpha$. In a similar vein, let y denote the realized number of

[50] To be more precise, if there is a type q (or rather $p(q)$) that exactly equals one of the equilibrium cutoffs, that type may be randomizing in equilibrium. For example, if i's signal is α and it turns out that $q_i = \bar{q}^a$, then i may randomize in equilibrium. However, such randomization will be accounted for in the pivotality calculus, and the formal equilibrium characterization is unaffected.

low quality news outlets (out of a total of m) such that $s_l = \alpha$. One thing to note is that given the realized x and y values, the conditional turnout rate is exactly the same in the two states. Moreover, the conditional *relative* turnout rates are exactly the same. To see why, let $P^j_{x,y,A}$ and $P^j_{x,y,B}$ denote the respective turnout rate for policy $j \in \{a, b\}$ conditional on x and y realizations in state A and state B. Given equilibrium cutoffs $\underline{q}^a$, $\bar{q}^a$, $\underline{q}^b$ and $\bar{q}^b$, $P^a_{x,y,A}$ equals

$$\frac{x}{n}\int_{\bar{q}^a}^{\bar{q}} qdF + \frac{y}{m}\int_{\bar{q}^a}^{\bar{q}}(1-q)dF + \frac{n-x}{n}\int_{\underline{q}}^{\underline{q}^a} qdF + \frac{m-y}{m}\int_{\underline{q}}^{\underline{q}^a}(1-q)dF.$$

It can be checked that $P^a_{x,y,B}$ is exactly the same; that is, $P^a_{x,y,A} = P^a_{x,y,B}$. This is true also with perception biases (with and without awareness). Therefore, hereafter $P^a_{x,y}$ denotes the turnout rate for policy a conditional on x and y realizations in either state. In a similar vein, it can be checked that $P^b_{x,y,A} = P^b_{x,y,B}$. Thus, hereafter $P^b_{x,y}$ denotes the turnout rate for policy b conditional on x and y realizations in either state. Finally let $P_{x,y}$ denote the total turnout rate conditional on x and y realizations in either state; that is, $P_{x,y} = P^a_{x,y} + P^b_{x,y}$. Note that the probability of selecting the correct policy equals

$$\sum_{x=0}^{n}\sum_{y=0}^{m}\sum_{T=0}^{N}\binom{N}{T}P^T_{x,y}(1-P_{x,y})^{N-T}(\pi M_A(x,y)P(a|x,y,T) + (1-\pi)M_B(x,y)P(b|x,y,T))$$

where $M_S(x,y)$ denotes the probability of (x,y) realizations in state $S \in \{A, B\}$, and $P(j|x,y,T)$ denotes the probability that policy $j \in \{a, b\}$ wins given x and y realizations and realized turnout being equal to T. For example, for odd T, $P(a|x,y,T)$ is equal to $\sum_{i=\frac{T+1}{2}}^{T}\binom{T}{i}\left(\frac{P^a_{x,y}}{P_{x,y}}\right)^i\left(\frac{P^b_{x,y}}{P_{x,y}}\right)^{T-i}$.

We can now prove the result below for unbiased electorates.

Lemma 9 *Let π be bounded above away from q_H and below away from $1-q_H$. In large elections with $q_H < 1$, optimal equilibria are such that the realization of x determines whether $\frac{P^a_{x,y}}{P_{x,y}} > \frac{1}{2}$ or $\frac{P^a_{x,y}}{P_{x,y}} < \frac{1}{2}$. More precisely, in every sufficiently large election, if $x \geq \frac{n+1}{2}$, then $\frac{P^a_{x,y}}{P_{x,y}} > \frac{1}{2}$, and if $x < \frac{n+1}{2}$, then $\frac{P^a_{x,y}}{P_{x,y}} < \frac{1}{2}$ (regardless of y). As a result, the probability of selecting the correct policy converges to $\sum_{i=\frac{n+1}{2}}^{n}\binom{n}{i}q_H^i(1-q_H)^{n-i}$ as $N \to \infty$. If $q_H = 1$, the probability converges to 1 as $N \to \infty$.*

Proof: The proof with $q_H = 1$ is straightforward given the strategy we constructed above. So, assume that $q_H < 1$. The probability of selecting the correct policy which we denote by C equals

$$\sum_{x=0}^{n}\sum_{y=0}^{m}\sum_{T=0}^{N}\binom{N}{T}P^T_{x,y}(1-P_{x,y})^{N-T}(\pi M_A(x,y)P(a|x,y,T) + (1-\pi)M_B(x,y)(1-P(a|x,y,T))$$

by what we have shown above. Note that for any $T \geq 0$ this term above reaches its highest possible value at $P(a|x, y, T) = 1$ if $\pi M_A(x, y) > (1-\pi) M_B(x, y)$ and at $P(a|x, y, T) = 0$ if $\pi M_A(x, y) < (1-\pi) M_B(x, y)$. Given that π is bounded away from q_H and $1 - q_H$, we have that $\pi M_A(x, y) > (1-\pi) M_B(x, y)$ if and only if $x \geq \frac{n+1}{2}$, and $\pi M_A(x, y) < (1-\pi) M_B(x, y)$ if and only if $x < \frac{n+1}{2}$ (recall than n is odd). As a result, C is bounded above by $\sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q_H^i (1-q_H)^{n-i}$. However, since the strategy that we constructed before generates a probability of selecting the correct policy which converges to $\sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q_H^i (1-q_H)^{n-i}$ as N goes to infinity, the probability in the optimal equilibrium must also converge to $\sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q_H^i (1-q_H)^{n-i}$. We can now prove the lemma. Suppose towards a contradiction that there exists an electorate size sequence for which the claim does not hold in the optimal equilibrium; e.g., there exists some $x \geq \frac{n+1}{2}$ such that $\frac{P_{x,y}^a}{P_{x,y}} \leq \frac{1}{2}$ for every element in that sequence. Then, it can be checked that the limiting probability of selecting the correct policy is bounded above away from $\sum_{i=\frac{n+1}{2}}^{n} \binom{n}{i} q_H^i (1-q_H)^{n-i}$, contradicting optimality. Hence, the lemma is proved.

We now consider a finite type space with $\{q_1, q_2, ..., q_T\}$ where $q_1 = \underline{q}$ and $q_T = \bar{q} \leq 1$ (note that none of the results we proved above rely on a continuum type space). We assume $q_H < 1$ to show the inefficiency caused by a high level of overconfidence. The proof with $q_H = 1$ is similar and much simpler (therefore, omitted). First, assume that $\limsup_{N\to\infty} \bar{q}_N^a \leq \limsup_{N\to\infty} \bar{q}_N^b$ without loss of generality. If there exists a type $\tilde{q} \leq \bar{q}$, such that $\limsup_{N\to\infty} \bar{q}_N^a < \tilde{q}$, then (abusing notation) there exists an electorate size sequence and cutoff sequence $(\bar{q}_N^a, \underline{q}_N^a, \bar{q}_N^b, \underline{q}_N^b)$ such that $\bar{q}_N^a < \tilde{q}$, $\bar{q}_N^b \geq \bar{q}_N^a$ and $\underline{q}_N^b < \tilde{q}$ for all N. Thus, $p_o^{-1}(\bar{q}_N^b) \geq p_o^{-1}(\bar{q}_N^a)$.[51] As a result, high enough levels of overconfidence implies that

$$\sum_{p_o^{-1}(\bar{q}_N^a) \leq q < \bar{q}_N^a} \lambda(q)(1-q)f(q) + \sum_{q > \bar{q}_N^a} (1-q)f(q) + \frac{n-1}{2n}\left(\sum_{p_o^{-1}(\bar{q}_N^a) \leq q < \bar{q}_N^a} \lambda(q) q f(q) + \sum_{q > \bar{q}_N^a} q f(q)\right)$$

$$\geq \sum_{q < p_o^{-1}(\underline{q}_N^b)} \left(1 - q + \frac{n-1}{2n} q\right) f(q) + \sum_{p_o^{-1}(\underline{q}_N^b) \leq q < \underline{q}_N^b} (1-\lambda(q))\left((1-q) + \frac{n-1}{2n} q\right) f(q) +$$
$$\frac{n+1}{2n}\left(\sum_{p_o^{-1}(\bar{q}_N^b) \leq q < \bar{q}_N^b} \lambda(q) q f(q) + \sum_{q > \bar{q}_N^b} q f(q)\right)$$

provided that the sufficient condition holds (we simplify notation by assuming $\underline{q}_N^a \leq \underline{q}$; if $\underline{q}_N^a > \underline{q}$, this only makes the inequality easier to obtain).[52] For example, with $p_o^{-1}(\tilde{q}) = \underline{q}$

[51] For the discrete case, $p_o^{-1}(\bar{q}_N^a) = \min\{q_t | p_o(q_t) \geq \bar{q}_N^a\}$.

[52] We assume without loss of generality that no type coincides with $\bar{q}_N^a$, $\bar{q}_N^b$, or $\underline{q}_N^b$. This is only to reduce notation and has no effect on the result.

(i.e., $p_o^{-1}(\bar{q}_N^a) = \underline{q} \leq p_o^{-1}(\bar{q}_N^b)$), and $\{\lambda(q)\}_{q \geq \underline{q}}$ sufficiently high for every $q \leq \tilde{q}$, the latter term will be weakly smaller than the former term due to the sufficient condition. Thus, if $x = \frac{n-1}{2}$, and $y = m$, then the inequality will imply that $\frac{P_{x,y}^b}{P_{x,y}} \leq \frac{1}{2}$, which is inefficient.[53] Next, assume that $\limsup_{N\to\infty} \bar{q}_N^a \geq \bar{q}$. This holds if and only if $\limsup_{N\to\infty} \bar{q}_N^b \geq \bar{q}$. In fact, these should be equalities as otherwise a violation of Lemma 9 follows (for example, $\limsup_{N\to\infty} \bar{q}_N^a > \bar{q}$ violates Lemma 9 if $x = n$ and $y = m$). Thus, $\limsup_{N\to\infty} \bar{q}_N^a = \limsup_{N\to\infty} \bar{q}_N^b = \bar{q}$. Let $\sigma_N(\bar{q}, s)$ denote the probability that an individual i with precision $q_i = \bar{q}$ and signal s votes for the policy that matches s with electorate size N and assume without loss of generality that $\sigma_N(\bar{q}, \alpha) \geq \sigma_N(\bar{q}, \beta)$ for all large N.[54] There are two possibilities to consider. First, consider the case where $\limsup_{N\to\infty} \sigma_N(\bar{q}, \alpha) = 0$. This implies that $\limsup_{N\to\infty} \sigma_N(\bar{q}, \beta) = 0$. Thus, (abusing notation) there exists an electorate size and cutoff sequence $(\bar{q}_N^a, \underline{q}_N^a, \bar{q}_N^b, \underline{q}_N^b)$ such that $\sigma_N(\bar{q}, \alpha) \to 0$ and $\sigma_N(\bar{q}, \beta) \to 0$. This requires $\underline{q}_N^a \leq \underline{q}$ and $\underline{q}_N^b \leq \underline{q}$ for all sufficiently large N. To see why, note that $\frac{P_{x,y}^b}{P_{x,y}} > \frac{1}{2}$ surely fails to hold if $\underline{q}_N^a > \underline{q}$, $x = 0$, and $y = 0$ given that $\sigma_N(\bar{q}, \beta) \to 0$. In fact, by the same argument, $\sigma_N(\underline{q}, \alpha) \to 0$ and $\sigma_N(\underline{q}, \beta) \to 0$, where $\sigma_N(\underline{q}, s)$ denotes the probability that an individual with $q_i = \underline{q}$ votes against s. In particular, $\sigma_N(\underline{q}, \alpha) f(\underline{q}) < \sigma_N(\bar{q}, \alpha) f(\bar{q})$ must hold due to Lemma 9 and thus, $\frac{\sigma_N(\bar{q},\alpha)}{\sigma_N(\underline{q},\alpha)}$ is bounded below away from 0 (consider $x = n$ and $y = m$). As a result, with $p_o^{-1}(\bar{q}) = \underline{q}$ and $\{\lambda(q)\}_{q \geq \underline{q}}$ sufficiently high, we find

$$\sigma_N(\bar{q}, \alpha) \left(\sum_{\underline{q} \leq q < \bar{q}} \lambda(q)(1-q) f(q) + (1-\bar{q}) f(\bar{q}) + \frac{n-1}{2n} \left(\sum_{\underline{q} \leq q < \bar{q}} \lambda(q) q f(q) + \bar{q} f(\bar{q}) \right) \right) +$$
$$\sigma_N(\underline{q}, \beta)(1 - \lambda(\underline{q})) f(\underline{q}) \underline{q} \frac{n+1}{2n}$$

to be weakly strictly greater than

$$\sigma_N(\bar{q}, \beta) \frac{n+1}{2n} \left(\sum_{\underline{q} \leq q < \bar{q}} \lambda(q) q f(q) + \bar{q} f(\bar{q})) \right) + \sigma_N(\underline{q}, \alpha) \left(1 - \underline{q} + \underline{q} \frac{n-1}{2n} \right) (1 - \lambda(\underline{q})) f(\underline{q})$$

provided that the sufficient condition holds. This implies that if $x = \frac{n-1}{2}$, and $y = m$, then $\frac{P_{x,y}^b}{P_{x,y}} > \frac{1}{2}$ fails to hold, violating Lemma 9. The proof for the case where $\limsup_{N\to\infty} \sigma_N(\bar{q}, \alpha) >$

[53]The proof is still valid if there exists an electorate size and cutoff subsequence $(\bar{q}_{N_k}^a, \underline{q}_{N_k}^a, \bar{q}_{N_k}^b, \underline{q}_{N_k}^b)$ such that $\bar{q}_{N_k}^b < \bar{q}_{N_k}^a < \tilde{q}$ for $N_k \to \infty$. In this case, the electorate is just as or even more prone to the ineficiency caused by overconfidence since $\limsup_{N_k\to\infty} \bar{q}_{N_k}^b \leq \limsup_{N\to\infty} \bar{q}_N^a$. In this scenario, the relevant case to consider is the case where $x = \frac{n+1}{2}$ and $y = 0$.

[54]The proof is similar if there exists an electorate size and cutoff subsequence $(\bar{q}_{N_k}^a, \underline{q}_{N_k}^a, \bar{q}_{N_k}^b, \underline{q}_{N_k}^b)$ such that $\bar{q}_{N_k}^b < \bar{q}_{N_k}^a < \tilde{q}$ for $N_k \to \infty$.

0 is analogous. The case with awareness can be analyzed in a similar way to previous extensions assuming that $q_H = 1$ and $\bar{q} = 1$, and considering the case where either $x = \frac{n-1}{2}$, and $y = m$ or $x = \frac{n+1}{2}$, and $y = 0$.

A.4 Proposition 5

Assume that π is bounded away from 0 and 1, and that $\bar{q} > 1 - \underline{q}$. If media veracity is low enough so that $\int_{\underline{q}}^{\bar{q}} q dF < 0.5$ holds, and sufficiently many individuals with low competence are highly overconfident and vote for their signal, then the probability that the wrong policy is chosen goes to one in at least one state as N grows without bound, whereas in an unbiased electorate the correct policy is chosen with a probability that goes to one in the optimal equilibrium.

Proof: The proof with unawareness is analogous to the proof of Proposition 3 with unawareness since this setting is in a way a special case where $m(q) = q$, and not only $\int_{\underline{q}}^{\bar{q}} m(q)dF < 0.5$ but also $\int_{\underline{q}}^{\bar{q}} q dF < 0.5$. The only difference is that we can prove general results even with awareness. Consider the case in which $\bar{q} < 1$. Consider those equilibria such that the correct policy is chosen with a probability that goes to one in both states. Let q_N^a and q_N^b denote the respective equilibrium cutoff representation for electorate size N as before, and uppose towards a contradiction that $\limsup_{N\to\infty} q_N^a \geq \bar{q}$ and $\limsup_{N\to\infty} q_N^b \geq \bar{q}$ (as in the proof of Proposition 3, $\limsup_{N\to\infty} q_N^a \geq \bar{q}$ if and only if $\limsup_{N\to\infty} q_N^b \geq \bar{q}$).[55] As in the proof of Proposition 3, a strict inequality cannot hold as that would mean that the probability of winning goes to either 0.5 or 0 for the correct policy in at least one state as $N \to \infty$. Hence, $\limsup_{N\to\infty} \frac{(1-\pi)\Pr(piv_a|B,N)}{\pi \Pr(piv_a|A,N)} = \frac{\bar{q}}{1-\bar{q}}$ and $\limsup_{N\to\infty} \frac{(1-\pi)\Pr(piv_b|B,N)}{\pi \Pr(piv_b|A,N)} = \frac{1-\bar{q}}{\bar{q}}$. Thus, (with an abuse of notation and using the notation introduced above) there exists an equilibrium sequence $q_N^a \to \bar{q}$ and $q_N^b \to \bar{q}$ giving rise to $\lim_{N\to\infty} \frac{(1-\pi)(x_N+y_N)}{\pi(w_N+z_N)} = \frac{\bar{q}}{1-\bar{q}}$ and $\lim_{N\to\infty} \frac{(1-\pi)(x_N+\gamma_N y_N)}{\pi(w_N+\xi_N z_N)} = \frac{1-\bar{q}}{\bar{q}}$. Thus, $\lim_{N\to\infty} \frac{x_N+y_N}{x_N+\gamma_N y_N} \frac{w_N+\xi_N z_N}{w_N+z_N} = \left(\frac{\bar{q}}{1-\bar{q}}\right)^2$. Note that $\gamma_N < 1 < \xi_N$ and thus,

$$\left(\frac{\bar{q}}{1-\bar{q}}\right)^2 = \lim_{N\to\infty} \frac{x_N + y_N}{x_N + \gamma_N y_N} \frac{w_N + \xi_N z_N}{w_N + z_N} < \limsup_{N\to\infty} \frac{\xi_N}{\gamma_N} \leq \left(\frac{\bar{q}}{1-\bar{q}}\right)^2$$

The strict inequality holds because Claim 2 above still holds in this case. We obviously get a contradiction with $\lim_{N\to\infty} \xi_N = 1 = \lim_{N\to\infty} \gamma_N$ (taking convergent subsequences if

[55] This holds under awareness of others' perception biases since we assume that $p_o(q) < \bar{q}$ for all $q < \bar{q}$ and $p_o(\bar{q}) = \bar{q}$. Thus, $p_o^{-1}(q) \to \bar{q}$ as $q \to \bar{q}$.

necessary); therefore, at least one of $\lim_{N\to\infty}\xi_N > 1$ or $\lim_{N\to\infty}\gamma_N < 1$ must hold. The weak inequality holds because

$$\gamma_N = \frac{\int_{q_N^a}^{\bar{q}}(1-q)dF + \int_{p_o^{-1}(q_N^a)}^{q_N^a}\lambda(q)(1-q)dF}{\int_{q_N^b}^{\bar{q}} qdF + \int_{p_o^{-1}(q_N^b)}^{q_N^b}\lambda(q)qdF} = \frac{1-\tilde{q}_N^a}{\tilde{q}_N^b}\frac{1-F(q_N^a) + \int_{p_o^{-1}(q_N^a)}^{q_N^a}\lambda(q)dF}{1-F(q_N^b) + \int_{p_o^{-1}(q_N^b)}^{q_N^b}\lambda(q)dF},$$

and

$$\xi_N = \frac{\int_{q_N^a}^{\bar{q}} qdF + \int_{p_o^{-1}(q_N^a)}^{q_N^a}\lambda(q)qdF}{\int_{q_N^b}^{\bar{q}}(1-q)dF + \int_{p_o^{-1}(q_N^b)}^{q_N^b}\lambda(q)(1-q)dF} = \frac{\tilde{q}_N^a}{1-\tilde{q}_N^b}\frac{1-F(q_N^a) + \int_{p_o^{-1}(q_N^a)}^{q_N^a}\lambda(q)dF}{1-F(q_N^b) + \int_{p_o^{-1}(q_N^b)}^{q_N^b}\lambda(q)dF},$$

where $\tilde{q}_N^a \in (p_o^{-1}(q_N^a), \bar{q})$ and $\tilde{q}_N^b \in (p_o^{-1}(q_N^b), \bar{q})$ are conditional expectations. Hence, we have a contradiction, and at least one of $\limsup_{N\to\infty} q_N^a < \bar{q}$ or $\limsup_{N\to\infty} q_N^b < \bar{q}$ must hold. In particular, either q_N^a or q_N^b is bounded above by a number $\tilde{q} < \bar{q}$ regardless of the form of overconfidence in the population. As a result, in at least one state, a sufficiently high level of overconfidence will prevent information aggregation and result in the wrong policy being chosen with a probability that goes to one. If $\bar{q} = 1$, then we assume for tractability (as we did in Proposition 3) that every i is an a-partisan with probability $\frac{p}{2}$ and a b-partisan with probability $\frac{p}{2}$. (The assumption that every i is an a-partisan with probability $\frac{p}{2}$ and a b-partisan with probability $\frac{p}{2}$ would not affect the proof with $\bar{q} < 1$ as it still implies that $\limsup_{N\to\infty}\frac{\xi_N}{\gamma_N} < \left(\frac{\bar{q}}{1-\bar{q}}\right)^2$.) In this case, $q_N^a \to 1$ and $q_N^b \to 1$ give rise to $\lim_{N\to\infty}\frac{x_N+y_N}{x_N+\gamma_N y_N}\frac{w_N+\xi_N z_N}{w_N+z_N} = \infty$. However, this is impossible because $\lim_{N\to\infty}\frac{x_N+y_N}{x_N+\gamma_N y_N}\frac{w_N+\xi_N z_N}{w_N+z_N} \leq \limsup_{N\to\infty}\frac{\xi_N}{\gamma_N} < \infty$ due to $p > 0$. Thus, at least one of $\limsup_{N\to\infty} q_N^a < 1$ or $\limsup_{N\to\infty} q_N^b < 1$ must hold. As a result, in at least one state, a sufficiently high level of overconfidence will prevent information aggregation and result in the wrong policy being chosen with a probability that goes to one.

A.5 Proofs of Examples in Section 2.3

Proof of Statements in Example 2: We first show that the equilibrium consists of cutoffs also in this variation. First, consider the case where the signal of individual i is α. In that case, i prefers voting for a over abstention if and only if

$$\frac{1}{2}\left(\Pr(piv_a \cap S = A|s_i = \alpha) - \Pr(piv_a \cap S = B|s_i = \alpha)\right) \geq 0$$

holds given as before. Different from the previous models, we must differentiate between the case where s_i comes from the high media veracity state ($v = h$) and the low media veracity state ($v = l$). It can be checked that $\Pr(piv_a \cap S = A | s_i = \alpha)$ equals $\Pr((piv_a \cap S = A \cap v = h) \cap (piv_a \cap S = A \cap v = l) | s_i = \alpha)$, and thus, $\Pr(piv_a \cap S = A | s_i = \alpha)$ equals

$$\frac{\Pr(piv_a \cap S = A \cap s_i = \alpha \cap v = h) + \Pr(piv_a \cap S = A \cap s_i = \alpha \cap v = l)}{\Pr(s_i = \alpha)},$$

where $\Pr(piv_a \cap S = A \cap s_i = \alpha \cap v = h)$ equals

$$\Pr(piv_a | S = A \cap v = h) \Pr(s_i = \alpha | S = A \cap v = h) \Pr(S = A | v = h) \Pr(v = h)$$

by conditional independence of individual signals. Thus, $\Pr(piv_a \cap S = A \cap s_i = \alpha \cap v = h)$ equals $\frac{\pi}{2} \Pr(piv_a | S = A \cap v = h) q$ where $q = \Pr(s_i = \alpha | S = A \cap v = h) = \Pr(s_i = \beta | S = B \cap v = h)$. Using a similar derivation for the case in which $v = l$, and letting $m(q) = \Pr(s_i = \alpha | S = A \cap v = l) = \Pr(s_i = \beta | S = B \cap v = l)$, we have that i weakly prefers voting for a over abstention and voting for b with $s_i = \alpha$ if and only if the term that represents the benefit of voting for a if $s_i = \alpha$, which is given by

$$\begin{aligned} \frac{\pi}{2}[q \Pr(piv_a | S &= A \cap v = h) + m(q) \Pr(piv_a | S = A \cap v = l)] - \\ \frac{(1-\pi)}{2}[(1-q) \Pr(piv_a | S &= B \cap v = h) + (1 - m(q)) \Pr(piv_a | S = B \cap v = l)] \end{aligned} \tag{10}$$

is weakly greater than both 0 and the term that represents the benefit of voting for b if $s_i = \alpha$, which is given by

$$\begin{aligned} \frac{(1-\pi)}{2}[(1-q) \Pr(piv_b | S &= B \cap v = h) + (1 - m(q)) \Pr(piv_b | S = B \cap v = l)] - \\ \frac{\pi}{2}[q \Pr(piv_b | S &= A \cap v = h) + m(q) \Pr(piv_b | S = A \cap v = l)]. \end{aligned} \tag{11}$$

Analogously, i weakly prefers voting for b over abstention and voting for a with $s_i = \alpha$ if and only if (11) is weakly greater than both 0 and (10). Next, consider the preference of i for voting for b over abstention and voting for a after observing a β signal: i weakly prefers voting for b over abstention and voting for a with $s_i = \beta$ if and only if

$$\begin{aligned} \frac{(1-\pi)}{2}[q \Pr(piv_b | S &= B \cap v = h) + m(q) \Pr(piv_b | S = B \cap v = l)] - \\ \frac{\pi}{2}[(1-q) \Pr(piv_b | S &= A \cap v = h) + (1 - m(q)) \Pr(piv_b | S = A \cap v = l)] \end{aligned} \tag{12}$$

is weakly greater than 0 as well as the term that represents the benefit of voting for a if $s_i = \beta$, which is given by

$$\begin{aligned} \frac{\pi}{2}[(1-q) \Pr(piv_a | S &= A \cap v = h) + (1 - m(q)) \Pr(piv_a | S = A \cap v = l)] - \\ \frac{(1-\pi)}{2}[q \Pr(piv_a | S &= B \cap v = h) + m(q) \Pr(piv_a | S = B \cap v = l)]. \end{aligned} \tag{13}$$

Analogously, i weakly prefers voting for a over abstention and voting for b with $s_i = \beta$ if and only if (13) is weakly greater than 0 as well as (11). Note that all these conditions give rise to a cutoff equilibrium structure as long as $f(q)$ is increasing in q (as in our example). This is because (10) and (12) are strictly increasing in q, whereas (11) and (13) are strictly decreasing in q if $m(q)$ is increasing in q as in our example. We will repeatedly use this cutoff structure in our proofs below. In particular, let $\nu_i^k(j)$ denote the turnout rate of type $k \in \{L, M, H\}$ for policy $i \in \{a, b\}$ having observed signal $j \in \{\alpha, \beta\}$. One can check that if for example $\nu_a^H(\alpha) \in (0, 1)$, then $\nu_a^M(\alpha) = \nu_a^L(\alpha) = 0$. Moreover, $\nu_a^H(\alpha) \in (0, 1)$ implies that $\nu_a^L(\beta) = \nu_a^M(\beta) = 0$ given the parameters of Example 2, because if $\nu_a^L(\beta) > 0$ or $\nu_a^M(\beta) > 0$, then $\nu_a^H(\alpha) = 1$ must hold, a contradiction. As another example, if $\nu_b^H(\alpha) > 0$ then $\nu_b^M(\alpha) = \nu_b^L(\alpha) = 1$, whereas if $\nu_b^L(\alpha) < 1$, then $\nu_b^M(\alpha) = \nu_b^H(\alpha) = 0$, and if $\nu_a^L(\beta) < 1$, then $\nu_a^M(\beta) = \nu_a^H(\beta) = 0$.

We now characterize equilibria further. Suppose in equilibrium $\nu_a^H(\beta) > 0$, meaning that H type votes against signal with strictly positive probability when $s = \beta$. Then, $\nu_a^M(\beta) = \nu_a^L(\beta) = 1$ and $\nu_a^H(\alpha) = 1$ must hold as argued above. Moreover, $\nu_b^M(\alpha) = \nu_b^L(\alpha) = 0$ since $\nu_b^H(\beta) < 1$. Next, note that we will assume away the case where $\nu_a^H(\alpha) = 0$ or $\nu_b^H(\beta) = 0$ in equilibrium as such equilibria are either unresponsive or inefficient and overconfidence can never improve the inefficiency.

We now show the negative impact of overconfidence in large elections when more than 2/3 of low type individuals are overconfident and perceive themselves as high type (with and without awareness). First, consider the case where $\nu_a^H(\alpha) \in (0, 1)$ in large elections. This implies that no L or M type votes for policy a regardless of signal. Regarding $\nu_b^H(\beta)$ there are two possibilities: either $\nu_b^H(\beta) \in (0, 1)$ with $\nu_a^H(\beta) = 0$ or $\nu_b^H(\beta) = 1$ (recall we already ruled out the case where $\nu_b^H(\beta) = 0$ and by what we showed above, $\nu_a^H(\beta) > 0$ requires $\nu_a^H(\alpha) = 1$). If $\nu_b^H(\beta) \in (0, 1)$, note that no L or M type votes for policy b regardless of signal; i.e., $\nu_b^L(j) = 0$ or $\nu_b^M(j) = 0$ for $j \in \{\alpha, \beta\}$. Now, assume without loss of generality that $\nu_b^H(\beta) \geq \nu_a^H(\alpha)$. It can be checked that if $v = l$ and more than 1/3 of low type individuals are overconfident and perceive themselves as high type, policy b will be chosen in state A with a probability that is very close to one, which is inefficient. Next, assume that $\nu_a^H(\alpha) < \nu_b^H(\beta) = 1$. This case now allows for the possibility that L and/or M types vote for policy b (but not for a) in equilibrium (in fact, even $\nu_b^H(\alpha) \in (0, 1)$ could be a possibility). But the conclusion is unchanged: if $v = l$ and more than a third of low type individuals are overconfident and perceive themselves as high type, policy b will be chosen in state A with a probability that is very close to one, which is inefficient. The case where

$\nu_b^H(\beta) < \nu_a^H(\alpha) = 1$ is analogous. Finally, we consider the case where $\nu_a^H(\alpha) = \nu_b^H(\beta) = 1$. In that case, 2/3 (or more) of low type individuals perceiving themselves as high type is sufficient for an inefficient election outcome. First, assume that all low type individuals vote against their signal with probability one; i.e., $\nu_a^L(\beta) = 1$ and $\nu_b^L(\alpha) = 1$. In fact, this may not be part of an equilibrium, but this is the largest possible barrier against the specific inefficiency that we are showing in at least one state of the world for $v = l$. Regarding the strategy of M type individuals, no matter which strategy they employ, if $v = l$ and more than 2/3 of low type individuals are overconfident and perceive themselves as high type, then the relative vote share for the correct policy is strictly lower than 0.5 in at least one state of the world given the parameters of the example. All of these statements hold regardless of whether or not individuals are aware of others' possible biases.

Proof of Statements in Example 4: We first show that we have an equilibrium if competent individuals (with $q_i = 1$) vote for the signal that matches their signal and the rest abstain. Modifying the equilibrium characterization in the proof of Proposition 4 for the case where $q_i \in \{0, 1\}$, it follows that for $q_i = 1$ and $s_i = \alpha$,

$$\pi \Pr(piv_a \cap s_i = \alpha | S = A \cap Q_i = H) > (1 - \pi) \Pr(piv_a \cap s_i = \alpha | S = B \cap Q_i = H)$$

must hold so that i strictly prefers voting for policy a over abstention, and

$$\pi \Pr(piv_b \cap s_i = \alpha | S = A \cap Q_i = H) > (1 - \pi) \Pr(piv_b \cap s_i = \alpha | S = B \cap Q_i = H)$$

so that i strictly prefers abstention over voting for policy b. Analogous inequalities must hold if $q_i = 1$ and $s_i = \beta$. For $q_i = 0$ and $s_i = \alpha$, given that $\pi = q_L = 0.5$, $\Pr(piv_a|S = A) < \Pr(piv_a|S = B)$, and $\Pr(piv_b|S = A) > \Pr(piv_b|S = B)$ must hold so that i strictly prefers abstention over voting for policy a or b. Analogously, $\Pr(piv_B|S = b) < \Pr(piv_B|S = a)$, and $\Pr(piv_a|S = B) > \Pr(piv_a|S = A)$ if $q_i = 0$ and $s_i = \beta$ (that $Q_i = L$ and $s_i = \alpha$ has no implication for piv_a since no j with $Q_j = L$ is supposed to vote). To show that the very first inequality holds, it is enough to show that $\frac{\Pr(piv_a \cap s_i = \alpha | S = A \cap Q_i = H)}{\Pr(piv_a \cap s_i = \alpha | S = B \cap Q_i = H)} > 1$ given that $q_H = 0.8$ and $\pi = 0.5$. Given the notation and results in the proof of Proposition 4 and our parameters, $\Pr(piv_a \cap s_i = \alpha | S = A \cap Q_i = H)$ equals

$$\sum_{x=1}^{5} M_A(x) \frac{x}{5} \sum_{T=0}^{N} \binom{N}{T} (0.25)^T (0.75)^{N-T} \binom{T}{\lfloor \frac{T}{2} \rfloor} \left(\frac{x}{5}\right)^{\lfloor \frac{T}{2} \rfloor} \left(\frac{5-x}{5}\right)^{T - \lfloor \frac{T}{2} \rfloor}$$

whereas $\Pr(piv_a \cap s_i = \alpha | S = B \cap Q_i = H)$ equals

$$\sum_{x=1}^{5} M_B(x) \frac{x}{5} \sum_{T=0}^{N} \binom{N}{T} (0.25)^T (0.75)^{N-T} \binom{T}{\lfloor \frac{T}{2} \rfloor} \left(\frac{x}{5}\right)^{\lfloor \frac{T}{2} \rfloor} \left(\frac{5-x}{5}\right)^{T - \lfloor \frac{T}{2} \rfloor}$$

where $M_A(x) = \binom{5}{x}(0.8)^x(0.2)^{5-x}$ and $M_B(x) = \binom{5}{x}(0.8)^{5-x}(0.2)^x$. For $x = 5$ it is immediate that the former term exceeds the latter term. Next, focusing separately on pairs $x \in \{2, 3\}$ and $x \in \{1, 4\}$ shows after some algebra that $\frac{\Pr(piv_a \cap s_i = \alpha | S = A \cap Q_i = H)}{\Pr(piv_a \cap s_i = \alpha | S = B \cap Q_i = H)} > 1$ must hold. The proof of other inequalities are similar.

We now prove that the ineffiency caused by overconfidence of 50 percent of individuals with $q_i = 0$ is around 16 percent. Assume that state is A. Given the parameters of our example, if $y = 0$ and $S = A$, then overconfident individuals' votes will result in the wrong policy being chosen with a probability that goes to one as N goes to infinity provided that $x \in \{3, 4, 5\}$, whereas if $x \in \{3, 4, 5\}$ in an unbiased electorate, then policy a is chosen with a probability that goes to one as N goes to infinity. This is due to the law of large numbers, and the fact that in However, if $y = 5$, then overconfident individuals' votes will result in the correct policy being chosen with a probability that goes to one as N goes to infinity if $x \in \{0, 1, 2\}$, whereas if $x \in \{0, 1, 2\}$ in an unbiased electorate, policy a is chosen with a probability that goes to zero as N goes to infinity. The net negative effect focusing on these scenarios with $y \in \{0, 5\}$ is ex ante equal to $\Pr(y = 0)\Pr(x \in \{3, 4, 5\}) - \Pr(y = 5)\Pr(x \in \{0, 1, 2\})$ which in turn equals $\Pr(y = 0)[\Pr(x \in \{3, 4, 5\}) - \Pr(x \in \{0, 1, 2\})] = 0.028$ since $q_L = 0.5$. If $y = 1$, then overconfident individuals' votes will result in the wrong policy b being chosen with a probability that goes to one as N goes to infinity if $x \in \{4, 5\}$ (whereas in an unbiased electorate, policy a is chosen with a probability that goes to one as N goes to infinity if $x \in \{3, 4\}$). However, if $y = 4$, then overconfident individuals' votes will result in the correct policy being chosen with a probability that goes to one as N goes to infinity if $x \in \{1, 2\}$. The net negative effect in these two scenarios with $y \in \{1, 4\}$ is ex ante equal to $\Pr(y = 1)[\Pr(x \in \{3, 4\}) - \Pr(x \in \{1, 2\})] = 0.048$. Finally, the net negative effect is equal to $\Pr(y = 2)[\Pr(x = 3) - \Pr(x = 2)] = 0.087$ if $y \in \{2, 3\}$. Summing up these numbers gives 16.3 percent. As the situation in state B is symmetric, we proved the result.

B Regression Analysis: Turnout at the Individual Level

We estimate a random effects probit regression model for a more detailed analysis of turnout behavior and to provide additional support for the rationality of voter behavior at the individual level. Specifically, we estimate a panel model explaining the individual decision whether or not to vote using the independent variables (i) elicited belief regarding own placement in the top 1/3; (ii) placement in top 1/3 (=1 if the subject is in top 1/3); (iii) elicited belief regarding other group members' likelihood of voting; and finally, (iv) time trend. Errors are

Table 2: Explaining Individual Turnout Decision

	(1) All Rounds	(2) Round=6
Placement in top 1/3	0.688 (0.239)***	-0.011 (0.172)
Elicited belief regarding own placement in top 1/3	0.028 (0.008)***	0.016 (0.006)***
Elicited belief regarding others' likelihood of voting	0.010 (0.005)**	0.014 (0.006)**
Elicited belief regarding (other) voters' placement in top 1/3	-	0.003 (0.006)
Round#	-0.026 (0.034)	-
Constant	-1.500 (0.631)**	-1.416 (0.778)*
Observations	864	144

Notes: The dependent variable is the subject's binary choice between voting (= 1) and abstaining (= 0). Errors are clustered at the session level. ***, **, and * indicates significance at the 1%, 5%, and 10% level, respectively.

clustered at the session level. Results are presented in Table 2. Coefficients of the variables "elicited belief regarding own placement in top 1/3" and "placement in top 1/3" are positive and highly significant ($p < 0.01$). These are consistent with the equilibrium prediction (as placement in top 1/3 is associated with higher elicited beliefs in the data). The coefficient of the variable "elicited belief regarding others' likelihood of voting" is also positive and significant ($p < 0.05$), whereas time trend is not significant. Thus, the regression analysis suggests that expectation of increased turnout from others encourages turnout. A theoretical mechanism behind this is as follows. Higher turnout from other members is associated with a lower cutoff used by them, and this implies a reduced expected signal precision for every other individual voter. If the subject weighs the negative effect of reduced expected precision of other voters more heavily than the positive effect of increased expected turnout of other voters, the subject's best response cutoff must decrease.

As we elicit subjects' beliefs regarding the competence of other voters only in the last round, we also run a regression using only the data of the last round and estimate a regression explaining the individual decision whether or not to vote in the last round using the independent variables (i) elicited belief about placement in the top 1/3; (ii) actual placement (whether or not the subject is in top 1/3); (iii) elicited belief regarding other group members' likelihood of voting; and (iv) elicited belief regarding other voters' placement in top 1/3.[56] Results and coefficients are largely consistent with the results of the regression above. However, the coefficient of the new variable and the coefficient of the variable "placement in top 1/3" are not significant, and their signs are not consistent with what we expect. This could be because data is very noisy since we use data from only one round in this regression, and very few people abstain in the last round (there is also correlation between two of our independent variables, and dropping one of them also does not change the outcome).

C Instructions

Experimental instructions are in English as the experiment was conducted in English. Instructions consist of five parts: four of them are paper based, and one is online. After the first paper-based part, subjects received online instructions for the quiz and the consequent belief elicitation task regarding quiz performance. We transcribe the online instructions below and present the screenshots of the quiz. In what follows, we reduced the font and the spacing of the original introductions to conserve space.

[56] We elicited subjects' beliefs regarding the probability that a randomly selected voter is in the top 1/3 only in the final round so as not to overburden them in every round with two belief elicitation tasks.

PART 1

Welcome to the experiment! Please turn off your cell phones and do not communicate with other subjects during the experiment. You will be paid for your participation in this experiment. The amount of money you earn depends on your decisions and decisions of other participants. Your decisions will be treated anonymously. The money you earn will be paid to you in cash at the end of the experiment.

This experiment consists of several parts. We explain the details of Part 1 now, the details of other parts will be explained in a short time.

Part 1 involves a "**Guessing Task**". In this task, you will be presented with various statements. Consider for example the statement "it snowed in Amsterdam in April 1991". We know whether or not this statement is true but you may not know it for certain. We will ask you to report your "best guess" about the chances that such a statement is true.

You will report your guess by choosing a percentage between 0 and 100. The percentage that you choose indicates your "best guess". **The higher the accuracy of your best guess, the higher the payoff you get.** In order to maximize your payoff:

- If you are certain a statement is true then you should choose a percentage of 100, and if you are certain a statement is false you should choose 0.
- In many cases you do not know for certain whether a statement is true or not. If you think the statement is equally likely to be true or false, you should choose a percentage of 50.
- **More generally, the more confident you are a statement is "true" the higher the percentage you should assign**. If for example you are *very confident* a statement is true, you should choose a percentage close to 100.
- **Conversely, the more confident you are a statement is "false" the lower the percentage you should assign**. If for example you are *very confident* a statement is false, you should choose a percentage close to 0.

You will make guesses regarding 6 statements in total, and 2 statements will be randomly selected in order to determine your payoff in this part. You will earn an amount from €0 to €1 in each selected statement depending on the accuracy of your guess.

You will earn the most if you "honestly report your best guess" about the chances that a statement is true because your payoff increases in the accuracy of your guess. (Your exact payoff for each selected statement is calculated as follows. Suppose you assign a percentage of P to the statement being true. If the statement is **true** your payoff equals $1-(1-P\%)^2$, and your payoff equals $1-(P\%)^2$ if the statement is **false.**)

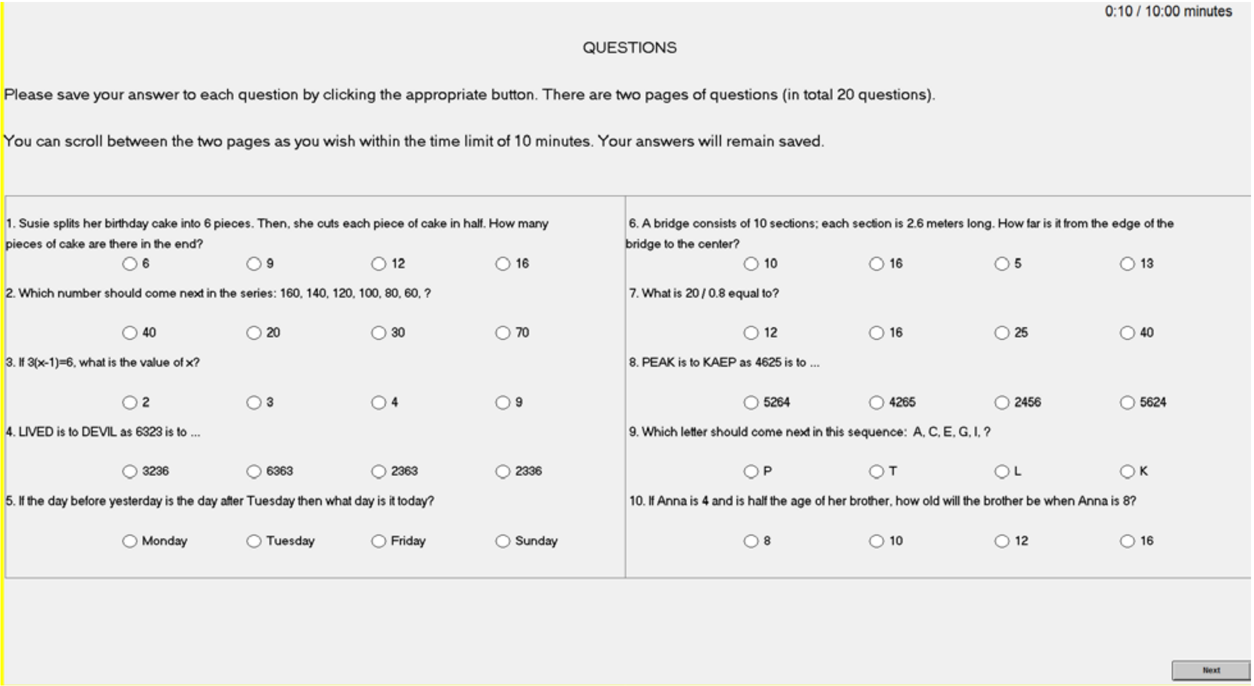
If you have any questions or need assistance of any kind please raise your hand and an experimenter will come to you. Please click OK on your screen when you are ready to start the experiment.

[*Part 1 was followed by the quiz and the subsequent belief elicitation task regarding quiz performance. The instructions for this part were only computerized. We transcribe the instruction screens below and present the screenshots of the quiz.*

Screen before the beginning of the quiz] In this part of the experiment, you will be taking a QUIZ on math and logic puzzles taken from various tests. The quiz involves **20 questions**. You will have **10 minutes** to correctly answer as many questions as you can. You will be paid **30 cents** for each correct answer. Your quiz score will also be relevant for later parts of the experiment. We will explain soon how exactly it will be relevant. Please click OK to continue.

[*Following screen before the beginning of the quiz*] You will next see the first page of the quiz. Please click the appropriate button to record your answer to a quiz question. When you want to see the second page of the quiz, click NEXT to continue. By clicking BACK on the second page you can go back to the first page of the quiz. You can go back and forth between the two pages as you wish within the time limit of 10 minutes. Answers that you have given will always remain saved when you move between the two pages. Please click START when you are ready to start the quiz.

[*Screenshots of the quiz: First screen*]



0:10 / 10:00 minutes

QUESTIONS

Please save your answer to each question by clicking the appropriate button. There are two pages of questions (in total 20 questions).

You can scroll between the two pages as you wish within the time limit of 10 minutes. Your answers will remain saved.

1. Susie splits her birthday cake into 6 pieces. Then, she cuts each piece of cake in half. How many pieces of cake are there in the end?
○ 6 ○ 9 ○ 12 ○ 16

2. Which number should come next in the series: 160, 140, 120, 100, 80, 60, ?
○ 40 ○ 20 ○ 30 ○ 70

3. If 3(x-1)=6, what is the value of x?
○ 2 ○ 3 ○ 4 ○ 9

4. LIVED is to DEVIL as 6323 is to ...
○ 3236 ○ 6363 ○ 2363 ○ 2336

5. If the day before yesterday is the day after Tuesday then what day is it today?
○ Monday ○ Tuesday ○ Friday ○ Sunday

6. A bridge consists of 10 sections; each section is 2.6 meters long. How far is it from the edge of the bridge to the center?
○ 10 ○ 16 ○ 5 ○ 13

7. What is 20 / 0.8 equal to?
○ 12 ○ 16 ○ 25 ○ 40

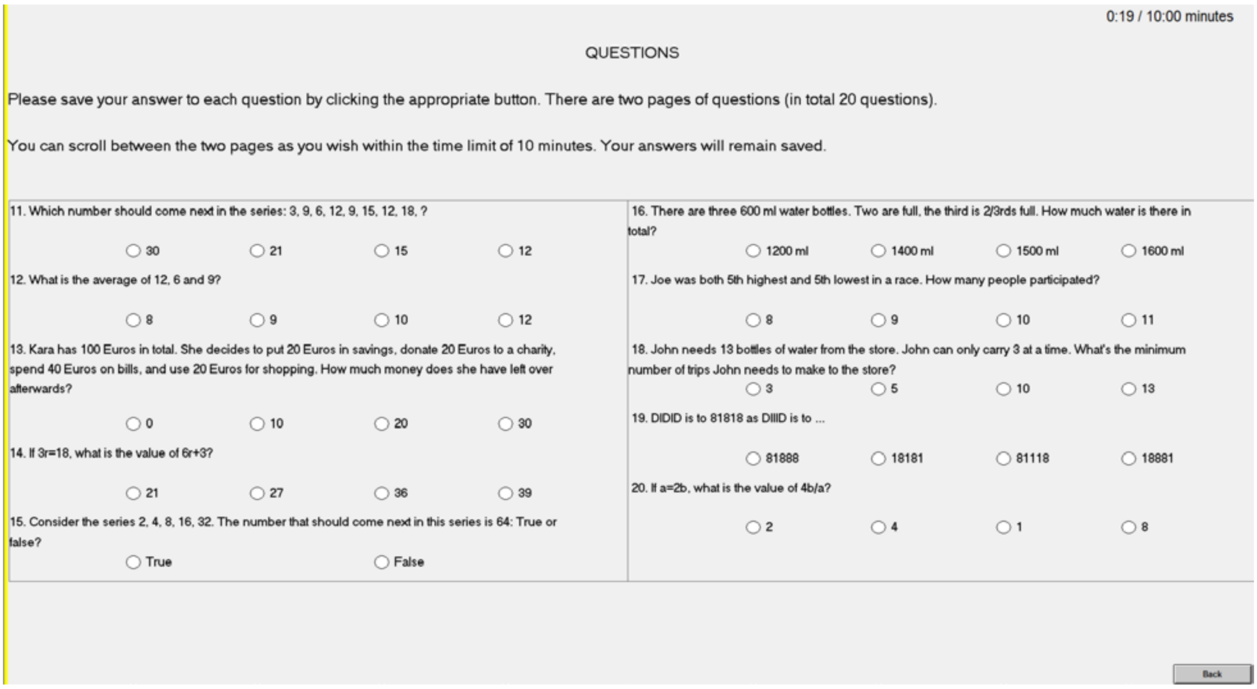
8. PEAK is to KAEP as 4625 is to ...
○ 5264 ○ 4265 ○ 2456 ○ 5624

9. Which letter should come next in this sequence: A, C, E, G, I, ?
○ P ○ T ○ L ○ K

10. If Anna is 4 and is half the age of her brother, how old will the brother be when Anna is 8?
○ 8 ○ 10 ○ 12 ○ 16

Next

[*Second quiz screen*]



0:19 / 10:00 minutes

QUESTIONS

Please save your answer to each question by clicking the appropriate button. There are two pages of questions (in total 20 questions).

You can scroll between the two pages as you wish within the time limit of 10 minutes. Your answers will remain saved.

11. Which number should come next in the series: 3, 9, 6, 12, 9, 15, 12, 18, ?
○ 30 ○ 21 ○ 15 ○ 12

12. What is the average of 12, 6 and 9?
○ 8 ○ 9 ○ 10 ○ 12

13. Kara has 100 Euros in total. She decides to put 20 Euros in savings, donate 20 Euros to a charity, spend 40 Euros on bills, and use 20 Euros for shopping. How much money does she have left over afterwards?
○ 0 ○ 10 ○ 20 ○ 30

14. If 3r=18, what is the value of 6r+3?
○ 21 ○ 27 ○ 36 ○ 39

15. Consider the series 2, 4, 8, 16, 32. The number that should come next in this series is 64: True or false?
○ True ○ False

16. There are three 600 ml water bottles. Two are full, the third is 2/3rds full. How much water is there in total?
○ 1200 ml ○ 1400 ml ○ 1500 ml ○ 1600 ml

17. Joe was both 5th highest and 5th lowest in a race. How many people participated?
○ 8 ○ 9 ○ 10 ○ 11

18. John needs 13 bottles of water from the store. John can only carry 3 at a time. What's the minimum number of trips John needs to make to the store?
○ 3 ○ 5 ○ 10 ○ 13

19. DIDID is to 81818 as DIIID is to ...
○ 81888 ○ 18181 ○ 81118 ○ 18881

20. If a=2b, what is the value of 4b/a?
○ 2 ○ 4 ○ 1 ○ 8

Back

[*Belief elicitation screen after the quiz ends*]

PLEASE READ CAREFULLY

We have now obtained the quiz scores of the 24 participants in this room, and ranked the participant scores from highest to lowest.

IMPORTANT: The quiz refers to the 20 math and logic puzzles that you have just answered, NOT the guessing tasks at the beginning of the experiment!

Your quiz score ranks in the **TOP 1/3 if at most 7 participants scored better than you in the quiz**. Exactly 8 out of 24 participants are in the top 1/3. For example, if you and another participant have the same quiz score and tie for the 8th place, then the tie is broken fairly and each of you is selected to be in the top 1/3 with equal chance.

We will now ask you to indicate your best guess about a statement regarding your score ranking. You will earn an amount from €0 to €3 depending on the accuracy of your guess. As before, you will indicate your guess choosing a percentage between 0 and 100, and as before, the higher the accuracy of your guess the higher the payoff you get; so you earn the most when you honestly report your best guess.

Please indicate your best guess about the statement below. What are the chances that it is true?

"My quiz score ranks in the top 1/3."

Please enter a percentage from 0 to 100 to indicate your best guess.

PART 2

*Part 2 and the following parts are on **group decision making**. From now on, you will be making choices in a **group**. In each round, the computer will randomly pick **RED** or **BLUE** as "**your group color**." You will not learn the color until the end of the round. **Your task as a group** is to try to **guess your group color correctly—based on information group members may receive**.*

Here is a detailed description of Part 2:

In each round, you will make choices in a **group of 24** (including you).

In each round, the computer will randomly pick either **RED** or **BLUE** as "**your group color**." There is a 50% chance **RED** will be picked and a 50% chance **BLUE** will be picked. In other words, **RED** and **BLUE** are *equally likely* to be your group color.

You will not learn your group color until the end of the round. Your task as a group is to try to guess the group color correctly. The group decision will be made by **voting**.

Before voting, each member of your group will be shown a "card", which may give information regarding your group color.

You will see only "**Your own Card"**. Similarly, each group member sees only his/her own card.

Each card is either red or blue. After you are shown **Your Card**, you will choose between voting for the color of Your Card and abstaining. In other words:

- If you are shown a **red** card then you choose between **voting for red** and **abstaining**.
- If you are shown a **blue** card then you choose between **voting for blue** and **abstaining.**

Cards are of two types: **informative** and **misleading**. A card is "**informative"** if its color is the **same** as your group color and it is "**misleading"** if it has the **opposite color**. In each round, Your Card is either informative or misleading.

*Since an "informative card" has the **same** color as your group, voting for the color of an informative card will result in a **CORRECT VOTE**. Since a "misleading card" has the **opposite** color, voting for the color of a misleading card will result in an **INCORRECT VOTE**.*

To repeat, Your Card is either **informative** or **misleading**. However, you **will NOT know** for certain whether Your Card is informative or misleading. This will be determined by **CHANCE** in each round. To be more precise, in each round, Your Card is

- an informative card with **X%** chance and a misleading card with **(100-X)%** chance.
- You will learn your X value before making your voting decision.
- At the beginning of every round, you will have a new X value that is randomly drawn from {1,2,3,…,99,100} by the computer. All possible values of X are equally likely.

Notice that:

- The closer X is to 100, the higher the chances that you have an informative card and observe your group's true color.
- The closer X is to 0, the higher the chances that you have a misleading card and observe the opposite color.

Here is an example: If your X value is exactly 50, then you are **equally likely to get an informative card as a misleading card.**

Another example: If your X value is 25, then you are **three times more likely to get a misleading card than an informative card,** and conversely, you are **three times more likely to get an informative card than a misleading card** if your X value is 75**.**

What about other members of your group? The rules for other members of your group are exactly the same as for you. Every member has his/her own X value that is randomly drawn from {1,2,3,…,99,100} by the computer. Every member observes his/her own Card, which is informative or misleading depending on the member's own X value. Note that you will NOT learn the X value or the card color of any other group member.

To summarize so far: After you learn your X value and the color of Your Card, you will choose between voting for the color of Your Card and abstaining. The same is true for every member of your group.

The color that receives a majority of the votes is the "group decision" and ties are broken fairly.[1] The group decision is "correct" if it is the same as your group color. You will earn **€4** if the group decision is correct and **€0** otherwise.

Reminder: When you observe Your Card's color, you will NOT know for certain whether or not Your Card is an informative card. However, you will know your X value, representing the chance with which Your Card is informative. Therefore, in your decision whether or not to vote, it is **important to weigh potential gains against potential losses GIVEN YOUR X VALUE.**

i. The more likely you are to have an informative card, the more likely you are to cast a correct vote and therefore the group decision is more likely to be correct if you vote. Hence, the higher the X value, the higher the **potential gains** from voting.

ii. Conversely, the more likely you are to have a misleading card, the more likely you are to cast a wrong vote and thus the group decision is more likely to be wrong if you vote. Hence, the lower the X value, the higher the **potential losses** from voting.

Therefore, if your X value is not sufficiently high, then **the potential loss** due to your voting is **higher** than **the potential gain** from your vote.

However, exactly which values of X allow your vote to generate higher potential gains than losses will depend on the behavior of other group members.

Because the precise value of X where the potential gains from your vote start dominating the potential losses depends on the voting behavior of other group members, we will also ask you to report your

[1] If Red and Blue receive the same number of votes then we will pick Red with 50% chance and Blue with 50% chance to determine the "group decision".

best guess about the chances that a randomly selected group member (other than you) chose to vote in each round.

As in the previous part, you will indicate your guess choosing a percentage between 0 and 100. As before, you earn the most if you honestly report your best guess. You will earn an amount from **€0** to **€1** for a guessing task depending on the accuracy of your guess.

You will play a total of **15 rounds** in this part and **2 rounds** will be randomly selected for payment. The amount you earn from the group decision and the guessing task in each of the selected rounds will be added to determine your payoff in this part. **Since these 2 rounds will be randomly selected, you should treat each round as a round you could be paid for.**

SUMMARY

In each round, you will make choices in a group of 24.

The computer will randomly pick **RED** or **BLUE** as your group color. Your task as a group is to try to guess your group color correctly. The group decision will be made by **voting**.

Each member of your group will be privately shown a "card". Each card is either red or blue.

You will only see the color of "**Your own Card"**.

Cards are of two types: **informative** and **misleading**. An "**informative"** card has the **same** color as your group, a "**misleading"** card has the **opposite color**.

You will NOT know for certain whether Your Card is informative or misleading. However, you will know your **X** value, representing the chance with which Your Card is informative.

After you observe the color of Your Card, you will choose between voting for the color of Your Card and abstaining.

The rules for other members of your group are *exactly the same as for you.*

The color that receives a majority of the votes is the "group decision." The group decision is "correct" if it is the same as your group color.

You will play a total of 15 rounds. 2 rounds will be randomly selected and the amount you earn from the group decision and the guessing task in the selected rounds will be added to determine your payoff in this part. Thus, you should treat each round as a round you could be paid for.

***

Please raise your hand if you have any questions. Please click OK when you are ready to start this part.

PART 3

This part is similar to Part 2. In Part 2, you did **NOT** know whether Your Card is informative or misleading with certainty as it was determined by **chance**. In this part, Your Card will be determined by **your quiz score** from Part 1, instead of being determined by chance.

You completed a quiz on math and logic puzzles in Part 1. As explained before, we ranked the quiz scores of the 24 participants in this room from highest to lowest. We know whether or not your quiz score is in the top 1/3 but you do **NOT** know it for certain.[2] In this part, Your Card will depend on your score as follows:

(1) If your quiz score is **in the top 1/3** then Your Card is an **informative card**

[2] Recall that exactly 8 out of 24 participants are in the top 1/3. If you and another participant tie for the 8th place, then each of you is selected to be in top 1/3 with equal chance.

(2) If your quiz score is **below the top 1/3** then Your Card is a **misleading card**[3]

After you took the quiz, we asked you to report your "best guess" about the chances that your quiz score ranks in the top 1/3—we will soon remind you of your guess. However, note that Your Card depends only on your true ranking, *not* on your guess.

What about other members of your group?

The rules for other members of your group are *exactly the same as for you*.

*To repeat, whether Your Card is informative or misleading depends on your quiz score. Thus, your belief regarding the chances that your quiz score ranks in the top 1/3 is analogous to your **X** value in Part 2.*

The group-decision making is the same as before. After you observe the color of Your Card, you will choose between voting for the color of Your Card and abstaining. The same is true for every member of your group. The color that receives a majority of the votes is the "group decision" and ties are broken randomly. The group decision is "correct" if it is the same as your group color. You will earn **€5** if the group decision is correct and **€0** otherwise.

Additionally, you will earn money in a Guessing Task just as in Part 2. In each round, you will report your best guess about the chances that a randomly selected member of your group (other than you) voted in that round. As before, you will indicate your guess choosing a percentage between 0 and 100. You can earn an amount from **€0** to **€1** in a guessing task depending on the accuracy of your guess.

You will play a total of **5 rounds** in this part. **1 round** will be randomly selected and the amount you earn from the group decision and the guessing task in that round will determine your payoff in this part.

Please raise your hand if you have any questions. Please click OK when you are ready to start this part.

PART 4

This is the final part of the experiment. This part is exactly the same as Part 3 **except that you will now also make a guess regarding the "competence of the average VOTER" in the room (other than you).** You will see the following information on your computer screen.

We have now randomly picked one member (other than you) that chose to **VOTE** in this round. Please indicate your best guess about the statement below: What are the chances it is true? "The quiz score of this randomly selected VOTER is in the top 1/3."

You will indicate your guess choosing a percentage between 0 and 100, as before. As before, you will also make a guess about the chances that a randomly selected member of your group (other than you) voted. As before, you will earn an amount from **€0** to **€1** in each guessing task depending on the accuracy of your guess.

You will play 1 **round** in this part. You will earn **€5** if the group decision is correct and **€0** otherwise. Please raise your hand if you have any questions. Please click OK to start.

[3] Recall that an informative card has the same color as your group, and a misleading card has the opposite color.