American Economic Review 2022, 112(10): 3367–3397
https://doi.org/10.1257/aer.20201844

Fake News, Voter Overconfidence, and the Quality of Democratic Choice[†]

By Melis Kartal and Jean-Robert Tyran*

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 then analyze richer models that allow for partisanship, targeted misinformation intended to sway public opinion, and news signals correlated across voters (due to media ownership concentration or censorship). In our experiment, overconfidence severely undermines information aggregation, suggesting that the effect of overconfidence can be much more pronounced at the collective than at the individual level. (*JEL* C91, D12, D72, D82, D83, L82)

Mass misinformation, now also known as *fake news*, is at the center stage of public and academic discourse because it disrupts the veracity of media coverage, which may affect public opinion and undermine the functioning of democracy (see, e.g., Allcott and Gentzkow 2017; Lazer et al. 2018). According to a survey study by the Pew Research Center in the United States, a sizable majority of respondents believe that fake news creates confusion about basic facts. However, 84 percent of these respondents are (somewhat or very) confident in their ability to recognize fake news.[1] According to another survey study, individuals believe that fake news has a greater effect on others than themselves (Jang and Kim 2018). These studies suggest *overconfidence* in one's ability to distinguish fact from fiction and that voters may be more prone to vote on the basis of misinformation than they think. Relatedly, a growing body of literature has focused on assessing the prevalence of misinformation and misperceptions; see Nyhan (2020) and the references therein. Our paper

*Kartal: Department of Economics, Vienna University of Economics and Business (email: melis.kartal@wu.ac.at); Tyran: Department of Economics, University of Vienna, Faculty of Business, Economics and Statistics; Department of Economics, University of Copenhagen, Faculty of Social Sciences (email: jean-robert.tyran@univie.ac.at). Stefano DellaVigna was the coeditor for this article. We are grateful for comments by participants of various conferences, seminars, and workshops, and would like to especially thank the coeditor, three anonymous referees, Amrita Dhillon, Sebastian Fehrler, Botond Kőszegi, Aniol Llorente-Saguer, Akaki Mamageishvili, César Martinelli, Salvatore Nunnari, Debraj Ray, Rupert Sausgruber, and James Tremewan for their very helpful comments and suggestions. We thank Christoph Rust for excellent research assistance. Our study was approved by the University of Vienna's Vienna Center for Experimental Economics (VCEE).

[†]Go to https://doi.org/10.1257/aer.20201844 to visit the article page for additional materials and author disclosure statements.

[1]A similar effect has been observed in an incentivized experiment by Serra-Garcia and Gneezy (2021). The authors show that subjects who observe videos in which senders either tell the truth or lie fail to detect lies and are overconfident in their ability to do so.

takes the natural next step and studies the effect of fake news and overconfidence on information aggregation and the quality of democratic choice.

Our theoretical analysis is based on the classic common-interest voting environment with two policy alternatives in which citizens share the objective of selecting the better policy but may differ in their information and hence disagree about which policy is better. Voting is costless, and abstention is possible as in Feddersen and Pesendorfer (1996). In contrast to the previous literature, we incorporate the possibility that individuals are overconfident (or underconfident) in their *competence* to obtain accurate news and form correct opinions. We analyze the effects of such biased perceptions on voting behavior and information aggregation, and we corroborate some theoretical predictions of our model in the laboratory. Our model shows that not only does overconfidence harm information aggregation, but perhaps more importantly, it can gravely exacerbate the impact of widespread misinformation on democratic decision-making. We discuss the results of several theoretical extensions. For example, we discuss the rationale behind the supply of misinformation and show how the interaction between overconfidence and misinformation can benefit a third party (or partisan voters when we relax the common-interest assumption).

Our theoretical framework is as follows. 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 outcome of the vote and the underlying state of the world, which is either *A* or *B*. Individuals have identical preferences and prefer the chosen policy to match the state of the world. In addition to a common prior regarding the state, each individual privately observes a binary signal. Signal precisions are heterogeneous as individuals differ in their competence. Hence, individuals who are less confident in their signal may rationally refrain from voting for their signal.

A key aspect of our model is that less competent individuals are more vulnerable to misinformation. Moreover, individuals hold subjective (and possibly inflated) views about their competence (e.g., they overestimate the quality of the news they receive and the accuracy of their opinions). Therefore, the rational mechanism in Feddersen and Pesendorfer (1996) in which less competent individuals refrain from following their signal may not work with voters being subject to the *Dunning-Kruger effect* (Kruger and Dunning 1999). According to this effect, incompetent individuals are prone to overconfidence and do not recognize their lack of competence. That is, they are "unskilled and unaware of it." Thus, our underlying hypothesis is that individuals who are more likely to be exposed to fake news, less able to discern correct and fake news, and less likely to form accurate opinions (i.e., "incompetent") are also more likely to be unaware and assess themselves as being competent. As a result, they are more likely to act on the basis of misinformation.[2]

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

[2] In a richer setup that allows for costly information acquisition, recognizing that one lacks competence on an issue may lead one to acquire information rather than abstain. Even so, the absence of overconfidence remains essential for informational efficiency (see also footnote 19).

Each individual privately observes their type, and this observation is either correct or overconfident. Assume a setting with high media veracity in which every competent individual obtains a perfectly informative signal and learns the true state, and every incompetent individual obtains a noisy signal that matches the true state with probability 0.7. The effect of overconfidence is very small in this setting as the average signal precision is quite high. Next, assume that media veracity is lower, making it more difficult for both types to infer the state correctly: a competent individual obtains a signal that matches the true state with probability 0.9, and an incompetent individual obtains an uninformative signal matching the state with probability 0.5. Comparing these two scenarios, the effect of overconfidence is much more pronounced when media veracity is lower.

As this example illustrates, overconfidence is typically more harmful at low levels of media veracity, acting as a "multiplier" of fake news. Hence, the powerful impact of the fake news phenomenon relies on the Dunning-Kruger effect in our model. More generally, we show that the quality of democratic decision-making depends on both the prevalence of misinformation in circulation and the extent of the Dunning-Kruger effect in the population.

We analyze the implications of fake news and overconfidence in richer models with correlated news signals, partisan voters, and asymmetric levels of media veracity in different states of the world. This asymmetry across different states may be caused by a third party, such as a special interest, corporate giant, or political agent, that strictly prefers a specific policy regardless of the state of the world and disseminates self-serving "news." Such news is benign in one state of the world, whereas in the other state it is misleading, giving rise to an asymmetry in the veracity of media coverage. We show that the third party can benefit from disseminating such news but *only* if the electorate is sufficiently overconfident.

To the best of our knowledge, we are the first to analyze a model with correlated signals in which several media outlets generate *statistically independent* news signals, and multiple individuals receive the same news signal from the same outlet. The ubiquitous assumption of the information aggregation literature that signals are independent across individuals 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 United States was 50 in 1983 and just 5 in 2004. We identify conditions under which overconfidence and the correlation of signals across voters jointly undermine information aggregation. Also, we note that such media concentration may facilitate third-party interference with media.[3]

To investigate the joint impact of overconfidence and misinformation, we conduct a series of experiments. In light of the discussion above, we expect high inefficiency in our *baseline* treatment, which involves a high level of misinformation and is conducive to overconfidence. We implement subjective beliefs regarding signal precision as follows. Subjects take an incentivized quiz with math and logic puzzles before (they learn about) the voting stage. They do not learn their quiz score

[3] 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, Marlow, Miller, and Roberts (2021) find that online bots have a substantial role in amplifying denialist messages about climate change as well as support for the then US President Trump's withdrawal from the Paris Agreement. See also our discussion in Section IC.

until the end of the experiment. In the voting stage, it is explained to subjects that the signal they get regarding the state depends on their quiz score; that is, a subject observes a signal that matches the true state (i.e., receives correct news) if and only if the subject's score places them in the top one-third of all the subjects in the same experimental session. Since subjects do not learn whether or not they are in the top one-third until the end of the experiment, they must form a belief on their likelihood of being in the top (i.e., a belief about the informativeness of their signal). We elicit subjects' beliefs in an incentivized manner. Our findings in the baseline reveal substantial overconfidence, which translates into excessive turnout and highly inefficient group decision-making consistent with our theory. These results indicate that collective overconfidence may result in more extreme outcomes than individual overconfidence: while collective decision-making can in theory cancel out modest levels of overconfidence, at high levels as in our experiment, it can result in drastic inefficiency. We check the robustness and consistency of our results in experimental variations of the baseline and show that limiting the extent of overconfidence or misinformation in the design improves collective decision-making as predicted.

Our theory and experiments suggest that pervasive overconfidence impairs information aggregation when a sufficiently high fraction of news is false. The obvious safeguards against this problem are to improve the reliability and veracity of the news media and to provide quality education for all. The news media has been dubbed the "fourth estate" because it has an essential role in maintaining checks and balances and limiting the power of special interests. For the media to fulfill this role, there must be sufficiently many high-quality news sources shielded from commercial or political interests trying to slant their reporting. This can potentially be achieved by generous public support for independent media (e.g., as in Finland and Germany), public service broadcasting, and regulations, such as network neutrality and caps on media ownership.[4] In addition, high-quality public education is essential to promote critical thinking skills and the competence to sort fact from fiction, which will minimize the fraction of unskilled and unaware voters vulnerable to manipulative news.

I. Model

An electorate of N citizens must choose one of two policies, a and b. Each individual votes for one of the two policies (at no cost) or abstains. The policy that receives a majority of the votes is chosen, with ties broken randomly. The utility of each individual depends on the chosen policy and an unknown state of the world, $S \in \{A,B\}$. The common prior is that the state is A with probability $\pi \in (0,1)$. Citizens have identical preferences and agree that policy a is superior in state A and policy b is superior in state B. In particular,

$$u(a|A) = u(b|B) = 1 \text{ and } u(a|B) = u(b|A) = 0. \tag{1}$$

[4] A wave of deregulations and commercialization has shaped the media in many countries since the 1980s. According to Bagdikian (2004, p. 139), the fairness doctrine in the United States required broadcast networks "to devote a reasonable time to discussions of serious public issues and allowed equal time for opposing views to be heard." The fairness doctrine was repealed in 1987, which arguably permitted the provision of news from partisan perspectives and transformed the media landscape in the following years.

Each individual i privately observes a noisy news signal, $s_i \in \{\alpha, \beta\}$, regarding the state. The term q_i denotes the precision of individual i's signal; that is, $q_i = \Pr(s_i = \alpha | S = A) = \Pr(s_i = \beta | S = B)$. For every i, q_i is private information and an i.i.d. draw from distribution $F[\underline{q}, \bar{q}]$. The distribution $F[\underline{q}, \bar{q}]$ is common knowledge and has strictly positive density f on $[\underline{q}, \bar{q}]$, where $\bar{q} > 0.5$. Our underlying hypothesis is that individuals differ in their competence and their ability to discern the true state especially because they differ in their cognitive (e.g., critical thinking and analytical reasoning) abilities. We interpret q_i as a measure of such competence.

Individual i is either *biased* or *unbiased* in their perception of q_i. More specifically, i perceives q_i as $p_i(q_i) \in [\underline{q}, \bar{q}]$, and $p_i(q_i) = q_i$ if i is unbiased, and $p_i(q_i) \neq q_i$ if i is biased. Biased individuals are either *overconfident* or *underconfident*. If i is overconfident, then $p_i(q_i) > q_i$, whereas $p_i(q_i) < q_i$ if i is underconfident.[5] Fixing q, $p_i(q) \equiv p_o(q)$ if i is overconfident, and $p_i(q) \equiv p_u(q)$ if i is underconfident, where $p_o(q)$ and $p_u(q)$ are continuous and increasing in q.[6] An individual with precision q is overconfident with probability $\lambda_o(q) > 0$, underconfident with probability $\lambda_u(q) \geq 0$, and otherwise unbiased, where $\lambda_o(q)$ and $\lambda_u(q)$ are continuous. The most relevant scenario (to be assumed later in Sections IB and IC) is the case where $\lambda_o(q)$ and $p_o(q)$ are high at low levels of q, i.e., the Dunning-Kruger effect.

A. *Equilibrium Analysis*

Individual i's strategy, σ_i, maps signal s_i and its subjective precision $p_i(q_i) \in \{p_o(q_i), q_i, p_u(q_i)\}$ to voting for a, voting for b, or abstention.[7] A profile of strategies $\{\sigma_i\}_{i \leq N}$ is a Bayesian Nash equilibrium (BNE) 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. In the main text, we assume that individuals are *unaware* of others' perception biases unless otherwise stated. Put differently, every i believes that every j is unbiased (i.e., $p_j(q_j) = q_j$) and acts on the basis of correct beliefs as would be the case in the standard model. All of our results in this section and many of our results in further sections are robust to awareness of others' perception biases (see, e.g., online Appendix A.2 and A.3). We focus on *symmetric* BNE in which all citizens choose ex ante the same strategy; that is, $\sigma_i = \sigma$ for all i. We characterize *responsive* equilibria. Unresponsive equilibria do not depend on s_i and are straightforward.

Every responsive BNE consists of "cutoffs." To see how equilibrium cutoffs arise, first assume that $s_i = \alpha$. In that case, i weakly prefers voting for policy a over abstention if and only if $\Pr(piv_a \cap S = A | s_i = \alpha)$ is weakly greater than $\Pr(piv_a \cap S = B | s_i = \alpha)$, where piv_a denotes the event in which i's vote for a is pivotal given σ_{-i}. By Bayes' rule and conditional independence, this is equivalent

[5] Individuals may derive "ego utility" from positive views about their abilities and competence (Kőszegi 2006). Overconfidence in q_i may be explained by such ego utility.

[6] Thus, $p_i(q)$ takes one of three possible values for fixed q: q, $p_o(q)$, and $p_u(q)$. This is without loss of generality in the sense that our results extend to settings with finitely many levels of overconfidence and underconfidence. See the last paragraph of the proof of Proposition 1 in online Appendix A.1.

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

to $p_i(q_i) \geq \Pr(S = B|piv_a)$. Hence, i compares the (perceived) precision of $s_i = \alpha$ to $\Pr(S = B|piv_a)$ in order to choose between voting for policy a and abstention. Also, i prefers voting for a over voting for b if and only if the *net* benefit of voting for a (i.e., $\Pr(piv_a \cap S = A|s_i = \alpha) - \Pr(piv_a \cap S = B|s_i = \alpha)$) is weakly greater than the *net* benefit of voting for b (i.e., $\Pr(piv_b \cap S = B|s_i = \alpha) - \Pr(piv_b \cap S = A|s_i = \alpha)$). This is equivalent to

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

by Bayes' rule and conditional independence. Combining these results, individual i with $s_i = \alpha$ prefers voting for a over other options if and only if $p_i(q_i) \geq q^a$, where

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

and $\eta = \frac{\Pr(piv_a)}{\Pr(piv_a) + \Pr(piv_b)}$. This also implies that if $s_i = \beta$, then i weakly prefers voting for a over other options if and only if $p_i(q_i) \leq 1 - q^a$. Symmetrically, if $s_i = \beta$, i weakly prefers voting for policy b over other options if and only if $p_i(q_i) \geq q^b$, where

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

and if $s_i = \alpha$, then i prefers voting for b over other options if and only if $p_i(q_i) \leq 1 - q^b$.

The cutoff conditions above involving s_i, $p_i(q_i)$, q^a, and q^b correspond to the following intuition: *an individual prefers voting for policy a (b) over other options if and only if the state is sufficiently likely to be A (B) based on the available information and $p_i(q_i)$, conditional on being pivotal.* Hence, we obtain the equilibrium characterization in Lemma 1 below.[8]

LEMMA 1: *A responsive BNE consists of two cutoffs q^a and q^b such that individual i votes (i) for policy a if and only if either $s_i = \alpha$ and $p_i(q_i) \geq q^a$ or $s_i = \beta$ and $p_i(q_i) \leq 1 - q^a$, and (ii) for policy b if and only if either $s_i = \beta$ and $p_i(q_i) \geq q^b$ or $s_i = \alpha$ and $p_i(q_i) \leq 1 - q^b$. If $\pi = 0.5$ or the correct policy is chosen in both states with a probability greater than* 0.5, *then $q^j = \Pr(S = B|piv_j)$ for $j \in \{a,b\}$. Furthermore, $q^a = q^b$ if $\pi = \underline{q} = 0.5$.*

Hereafter, we focus on the effect of overconfidence without loss of generality.[9] It is not possible to show the uniqueness of the equilibrium characterized in Lemma 1. Therefore, our efficiency benchmark is the *optimal unbiased equilibrium*, which

[8] Lemma 1 is general and applies to electorates with and without perception biases (with and without awareness in the former case). Also, the equilibrium of an electorate that consists of only unbiased individuals is *always* an equilibrium with perception biases and unawareness thereof.

[9] All of our results are robust to underconfidence; it typically exacerbates the effect of overconfidence.

maximizes the expected probability of selecting the correct policy in an electorate consisting of unbiased citizens.[10]

We first characterize sufficient conditions under which the optimal BNE is responsive and involves at least one interior cutoff; that is, at least one of q^a, q^b, $1-q^a$, and $1-q^b$ is an element of $(\underline{q},\bar{q})$. To see why this matters, consider the following example. If $\pi = \underline{q} = 0.5$ with N odd, the case where every i votes for the policy that matches s_i regardless of q_i is an equilibrium (i.e., $q^a = q^b = \underline{q}$). If this is the optimal unbiased equilibrium, then overconfidence has no impact on efficiency because it has no effect on equilibrium behavior. However, Lemma 2 shows that this case can be considered a knife-edge situation. Put differently, for every $\underline{q} \geq 0$, there exists a sizable set of π values such that the optimal unbiased equilibrium involves at least one interior cutoff.

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

Proposition 1 shows that under the parameter conditions specified in Lemma 2, overconfidence results in a deviation from the optimal unbiased equilibrium strategy, which is 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. However, under the conditions spelled out in Lemma 2, the equilibrium behavior of overconfident individuals is not a "true" best response. 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, the expected probability of selecting the correct policy is strictly lower relative to that in the optimal unbiased equilibrium. Proposition 1 substantially draws on the insight in McLennan (1998), but showing the existence of the optimal unbiased strategy and characterizing conditions under which overconfidence is harmful is challenging in our general setup.

PROPOSITION 1 (Inefficiency of Overconfidence): *Under the conditions on* $\underline{q}$ *and* π *specified in Lemma 2, the decision-making accuracy of the optimal unbiased equilibrium cannot be attained in an electorate with overconfident citizens.*

In general, $\underline{q}$ and $F[\underline{q},\bar{q}]$ depend on both the competence of the electorate and the veracity of the media. We are especially interested in the *interaction* of overconfidence and $F[\underline{q},\bar{q}]$ with $\underline{q} < 0.5$, which we associate with widespread

[10] Pareto optimality is commonly used as an equilibrium selection criterion in studies of voting and information aggregation. The optimal equilibrium strategy is particularly relevant for common interest voting games since it maximizes the expected payoff among *all* strategies as discussed below. See McLennan (1998) for a discussion of other desirable properties of the optimal equilibrium in common interest voting games.

misinformation and fake news. Due to fake news and overconfidence, $q_i < 0.5$ and $p_i(q_i) > 0.5$ may hold. That is, individuals who are highly susceptible to fake news may also be unaware of the fictitious nature of their signal.[11] They may take such news at face value and act upon it. Such a combination of fake news and overconfidence is potentially harmful. For example, it may lead to a case where the incorrect policy is more likely to be implemented than the correct policy in at least one state of the world, which is further *amplified* in larger elections. This point is important for Proposition 2 below, which informs our experimental design described in Section II.

B. *Large Elections*

The set of π values for which Proposition 1 holds grows in large elections. Nevertheless, the negative impact of overconfidence will also vanish if $\underline{q} \geq 0.5$. In other words, overconfidence can have a nontrivial effect in large elections only if $\underline{q} < 0.5$. In fact, the impact of overconfidence on large elections depends not only on $\underline{q}$ but also on $E(q)$ and the extent of overconfidence in the population. In particular, the share of misinformation in circulation must be high enough so that $E(q)$ is lower than 0.5, and the Dunning-Kruger effect must be sufficiently pervasive for overconfidence to have a significant impact in large elections. Intuitively, these conditions correspond to an excessive share of people who are more likely to be wrong than right in their opinions and their votes due to misinformation and overconfidence. These conditions are admittedly strong, but if they hold, the result is severe.

PROPOSITION 2 (Inefficiency of Overconfidence and Low Media Veracity in Large Elections): *In the limit as N goes to infinity, an unbiased electorate makes the correct decision with certainty, whereas a sufficiently overconfident electorate chooses the wrong policy in at least one state if* $E(q) < 0.5$.[12]

Proposition 2 relates to the profound concern about *disinformation warfare* since the United Kingdom's Brexit referendum and the 2016 US presidential election. In 2017, the then Prime Minister of the United Kingdom Theresa May accused Russia of "planting fake stories" in order to "sow discord in the West." According to the European Commission, Russia carried out a widespread misinformation campaign to influence the 2019 European Parliament election. In addition, extreme distrust of mainstream sources, antiscience views, and the growing tendency for conspiratorial thinking may be important sources of incompetence and untruths.[13] Thus, $E(q) < 0.5$ may hold in practice.[14] Indeed, a widely publicized study by Vosoughi,

[11] See online Appendix A.5 for a richer formulation of our model with $\underline{q} < 0.5$ where s_i represents the direction of i's opinion, and $F[\underline{q}, \bar{q}]$ represents the distribution of opinion precisions. Hence, $q_i < 0.5$ can hold only if i is overconfident. This model generates analogous theoretical findings but is more cumbersome.

[12] See online Appendix A.3 for the formal restatement and proof of Proposition 2.

[13] The psychology literature shows that conspiratorial thinking and low competence are associated (see, e.g., Swami et al. 2014; Ståhl and Van Prooijen 2018; and Lantian et al. 2021).

[14] There are large collective environments in which overconfidence may be harmful even if $E(q) \geq 0.5$, such as disease control. Vaccinations make it possible to reach herd immunity but only if sufficiently many individuals (typically, a large *supermajority*) are vaccinated. Misinformation and overconfidence may harm disease control even if $E(q) \geq 0.5$ by suppressing vaccine uptake and keeping the vaccinated share below the required supermajority.

Roy, and Aral (2018) shows that fake news reached more people and spread six times faster than the truth on social media. Proposition 2 also underlies the *key* aspects of our experimental design: high levels of misinformation and subjective beliefs that are often overconfident. In particular, our baseline experiment mimics the conditions of Proposition 2 with a design that implements $E(q) < 0.5$ and is conducive to substantial overconfidence.

C. *Extensions for Large Elections*

Can a version of Proposition 2 still obtain in a setup where *every* citizen is ex ante more likely to be correctly informed than misinformed? So far, our model has assumed the following:

(i) Citizens have identical preferences (i.e., there are no partisans).

(ii) The veracity of news does not depend on the state of the world; i.e., $\Pr(s_i = \alpha | S = A) = \Pr(s_i = \beta | S = B)$.

(iii) News signals (s_i) are conditionally independent across individuals.

Assumptions (i)–(iii) are standard in the literature but also strong. Below, we relax them one by one and show that overconfidence may harm information aggregation in large elections *even if every individual is ex ante more likely to be correctly informed than misinformed.*

(**i**) We start by allowing for the presence of partisans who *always* prefer and vote for a certain policy regardless of the state. Each individual is an a-partisan with probability p_a, a b-partisan with probability p_b, and otherwise a nonpartisan with preferences given by (1), where $p_a \leq p_b$ without loss of generality. We assume that $p_b < 0.5$ so that information aggregation is essential for efficiency. Proposition 3 shows that even when $\underline{q} \geq 0.5$, the joint force of partisanship and overconfidence among nonpartisans may hinder information aggregation. We restate Proposition 3 formally and prove it in online Appendix A.4.1.

PROPOSITION 3 (A sufficiently large partisan group may benefit from low media veracity and high voter overconfidence): *Assume that* $p_a < p_b < 0.5$ *and* $\underline{q} \geq 0.5$. *While an unbiased electorate makes the correct decision with a probability that goes to one in both states as N goes to infinity, a sufficiently overconfident electorate will choose policy b with a probability that goes to one in both states if* $E(q)$ *is sufficiently low.*

The intuition is as follows. Due to the asymmetry in the size of a- and b-partisans, implementing the correct policy in state A requires a *supermajority* of nonpartisans to vote for policy a.[15] In particular, if $E(q)$ is sufficiently low, individuals with low q

[15] Voter suppression, such as the laws introduced in the United States after the 2020 presidential election, may influence this asymmetry by changing the relative group size of nonpartisans as well as a- and b-partisans. Hence,

should either vote for a or abstain to counteract the excessive amount of b-partisan votes in state A. However, under a large-scale Dunning-Kruger effect, individuals with low q and β signals will largely follow their signals, which in turn harms information aggregation in state A.

(**ii**) We next relax the assumption that the distribution of q is identical in the two states. In particular, one state is associated with low media veracity and the other with higher media veracity. This scenario captures misinformation dissemination by a third party with a large stake in the policy outcome, such as a special interest, corporate lobby, or political entity, that strictly prefers a certain alternative irrespective of the state of the world.[16] The dissemination may take the form of media capture or public opinion manipulation via political propaganda, sponsored or paid articles, hidden advertisements, and expert testimonies, which we summarily denote by "sponsored news." Such news is innocuous in one state of the world, but false and misleading in the other. Sponsored news therefore increases obfuscation and reduces media veracity in only one state; i.e., the state in which the preferred policy of the third party is inferior. Low media veracity of that state may enable the third party to have its preferred policy implemented in both states, but *only* in sufficiently overconfident electorates. The simple example below illustrates this point.

Example 1: For simplicity, we assume that there are two types of individuals in terms of competence: H and 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 $p \in \{0, 5.1\}$. An L-type individual i is overconfident if $p_i(L) = H$. First, let $p = 0.5$. In this case, the correct policy is chosen with certainty in the limit—unless every H type is underconfident and abstains. What if a special interest that has a large stake in having policy a implemented can increase p from 0.5 to 1 (e.g., by inundating media outlets targeted at low types with sponsored news that promote policy a)? As we argue below, the special interest can indeed benefit from increasing p to 1 but only if the electorate is sufficiently overconfident. Importantly, an L type is *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. If $p = 0.5$, then roughly equal numbers of L-type individuals observe α and β signals in large elections, whereas if $p = 1$, then *all* L types observe an α signal. In particular, if $p = 1$, then all overconfident L

it may influence the likelihood of an inefficient outcome in the next election.

[16] For example, politicians (especially incumbents) resort to propaganda to sway opinions and voter behavior. As another example, various lobbies and industries have extensively used experts, think tanks, and media channels to disseminate self-serving news in an effort to influence public opinion and policymakers. See Kartal and Tremewan (2018) and the references therein for examples.

types vote *as if they were a-partisans*. Hence, if there are sufficiently many L types sufficiently many of whom are overconfident, policy a is chosen in both states with certainty in the limit as N goes to infinity, as desired by the special interest.

To characterize equilibria in a formal yet tractable setting, we model the signal precision of individual i as a function of a media veracity indicator in state $S \in \{A, B\}$, denoted by v_S, and individual competence q_i, which is an i.i.d. draw from distribution $F[\underline{q}, \bar{q}]$ as before. As v_S depends on the state of the world, the signal precision for individual i is drawn from related but different distributions in A and B (see the formal model and equilibrium characterization in online Appendix A.4.2). The difference in v_S across the two states is due to sponsored news disseminated by a third party that always prefers policy a. If the third party pursues a sufficiently large news campaign, the average signal precision in state B may fall below 0.5, implying that a majority will observe an α signal in state B. In that case, efficiency requires individuals with sufficiently low q to vote for b or abstain. However, under a sufficiently large-scale Dunning-Kruger effect, too many individuals will vote according to their α signals, and policy a will be chosen in *both* states with a probability that goes to one, as desired by the "source" of sponsored news. This will be the case even if every individual is ex ante more likely to be correctly informed than misinformed on average across the two states.

PROPOSITION 4 (A third party, such as a special interest, lobby, or political entity, may benefit from low media veracity and high voter overconfidence): *If media veracity in state B is sufficiently low* (*due to sponsored news*), *and the electorate is sufficiently overconfident*, *then policy a is chosen with certainty in both states in the limit as N goes to infinity. The unbiased electorate however chooses the correct policy in both states.*[17]

(**iii**) We finally relax the assumption that each individual observes a statistically independent signal. Instead, there are finitely many news sources generating statistically independent news signals, and each individual obtains one signal from one source. As a result, in large elections many individuals receive the same signal from the same news source. News sources differ in quality. There are $n \geq 1$ high-quality news sources and $m \geq 1$ low-quality sources. A high-quality source j provides a statistically independent signal $s_{Hj} \in \{\alpha, \beta\}$ with precision $q_H > 0.5$; i.e., $\Pr(s_{Hj} = \alpha | A) = \Pr(s_{Hj} = \beta | B) = q_H > 0.5$. A low-quality source k provides a statistically independent signal $s_{Lk} \in \{\alpha, \beta\}$ with precision $q_L = 0.5$; i.e., $\Pr(s_{Lk} = \alpha | A) = \Pr(s_{Lk} = \beta | B) = 0.5$. Individual i receives s_i from a high-quality (low-quality) news source with probability q_i (probability $1 - q_i$) with each high-quality (low-quality) source equally likely to provide s_i in that case. As before, q_i is an i.i.d. draw from distribution $F[\underline{q}, \bar{q}]$. Possible interpretations of correlated signals are various forms of media capture (or censorship) and high media ownership concentration observed in many countries.

Importantly, $q_i q_H + (1 - q_i) q_L \geq 0.5$ for every $q_i \geq 0$. Hence, every i is more likely to be correctly informed than misinformed in both states. In the

[17] See the formal restatement and proof of Proposition 4 in online Appendix A.4.2.

independent-signal model, this would be sufficient to ensure that the correct policy is chosen with certainty in the limit in both states regardless of the extent of overconfidence. However, the correlation among s_i's results in correlated and thus possibly large mistakes due to overconfidence. To see the intuition, consider an extremely simple example in which there is one outlet from each type of news source; that is, $n = m = 1$. Assume that there are two types of individuals: low type with $q_i = 0$ and high type with $q_i = 1$. In each state, *all* low types will observe the wrong signal with probability 0.5. In such a case, the inferior policy will be implemented if there are sufficiently many low-type individuals and sufficiently many of them are overconfident (i.e., $p(q_i) = 1$ for them). Following the same intuition, an inefficiency result obtains more generally when $n \geq 1$ and $m \geq 1$. That is, the negative impact of overconfidence on information aggregation may not vanish in the limit as N goes to infinity even though every individual is more likely to be correctly informed than misinformed.[18]

PROPOSITION 5 (Media Concentration with High Voter Overconfidence): *In the limit as N goes to infinity, the decision-making accuracy of the optimal unbiased equilibrium cannot be attained in a sufficiently overconfident electorate if $E(q)$ is sufficiently low.*

It is natural to combine the last two models involving media concentration and sponsored news disseminated by a third party. The formal extension is outside the scope of the current paper, but an important direction for future research since accessing and influencing the news media may be easier for third parties if a few conglomerates control most of the media. On a related note, it may be argued that social media alleviates the issue of the scarcity of statistically independent news signals relevant for Proposition 5. However, social media does not seem to be the democratizing force it was once thought to be. To the contrary, it may be an essential tool for spreading sponsored news and misinformation. For example, a widely publicized report states that a small number of "influencers" account for the majority of vaccine misinformation on social media in the United States.[19] Many of these influencers share "news" to promote their businesses to followers, which relates to the above-mentioned extension for future research.

We conclude the theory section with the following remarks. Turnout alone may not be a sufficient safeguard against problems caused by widespread misinformation and overconfidence. Therefore, it is essential to establish and strengthen institutions that support free and independent media and foster a competent, engaged, and ethical citizenry. The role of institutions relates to our model as follows. Supporting independent, investigative journalism and limiting media concentration improves $F[\underline{q}, \bar{q}]$ as well as v_S and n in the second and third extensions, respectively. A high-quality and ethical education not only improves competence q_i but also reduces

[18] See the equilibrium characterization of this model and the formal restatement of Proposition 5 along with its proof in online Appendix A.4.3.

[19] For a discussion of the report, "The Disinformation Dozen," by the Center for Countering Digital Hate, see Frenkel, Sheera. 2021. "The Most Influential Spreader of Coronavirus Misinformation Online." *New York Times* https://www.nytimes.com/2021/07/24/technology/joseph-mercola-coronavirus-misinformation-online.html.

$p(q_i) - q_i$ and $\lambda_o(q_i)$; that is, it improves self-awareness. This self-awareness is a virtue important for any form of civic and political participation.[20]

II. Experiment

The experiments reported below serve to test some of the implications of our theory (in particular, Propositions 1 and 2) in environments with widespread misinformation and overconfidence. Our experimental results also relate to the literature on individual overconfidence by showing that the outcome of collective decisions with overconfidence can be more extreme than the aggregate of individual decisions with overconfidence. That is, while collective decision-making may mitigate or balance out the overconfidence effect, we show here that it may also generate very inefficient outcomes in settings with high levels of fake news and overconfidence.[21]

A. *Experimental Design and Predictions*

We conduct four experiments. The voting treatment in each experiment is based on the game analyzed in Section I. The state of the world (described as "group color" in the experimental instructions) is either *Red* or *Blue*. Each state is ex ante equally likely (i.e., $\pi = 0.5$). There are two alternatives to vote on: *red* and *blue*. After observing a private signal, which is also either *red* or *blue*, each subject decides whether or not to vote. The group decision is determined by majority voting, and group members receive a reward if and only if the outcome of the vote matches the group color. Subjects make choices in relatively large lab electorates.

The individual signal precision q_i, its subjective perception $p_i(q_i)$, correct news, and fake news are implemented as follows. Subjects take a quiz on math and logic puzzles before learning about the voting phase. The quiz is incentivized as each correct answer is rewarded. Subjects do not learn about their performance in the quiz until the end of the experiment. The voting part in each experiment begins with a *control* condition followed by the treatment. In the treatment, subjects are informed that the signal they receive regarding the group color, s_i, 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 them in the top $1/x$ of all the subjects in the same experimental session, where $x \in \{2, 3\}$ depends on the experiment. Hence, signal precision q_i and its subjective perception $p_i(q_i)$ depend on the (perceived) ability to perform well in the quiz. Since subjects do not learn about their quiz performance and ranking until the end of the experiment, they must form a belief, $p_i(q_i)$, regarding the likelihood of scoring in the top $1/x$ of all the subjects in the same session. We elicit beliefs after the quiz to obtain a measure of $p_i(q_i)$.

[20] We focus only on voting, but the choice between voting and abstention relates to or can be combined in future work with various forms of civic (non)participation. Under the common preference assumption, decisions concerning protest participation, information acquisition, communication and persuasion, donations, and even running for office rely on how informed and competent an individual perceives themselves to be. Therefore, they are subject to similar misinformation and overconfidence problems as in our model.

[21] See Fehr and Tyran (2005) for a general discussion of how individual biases and aggregate outcomes may relate.

We use relative performance to determine the signal of a subject for three reasons. First, the theoretical mechanism of the "swing voter" model (Feddersen and Pesendorfer 1996) concerns relative competence, not absolute competence. Second, it enables us to use the techniques developed in Benoît and Dubra (2011) and Benoît, Dubra, and Moore (2015) in order to test for *true* overconfidence. Third, it allows us to fully control the share of misinformation in a laboratory election setting and avoid noise in that respect.

Our experiments vary x (i.e., the prevalence of fake news), the difficulty of the quiz (i.e., the level of overconfidence among subjects), and "informative voting" as defined below. In particular, our study consists of a *baseline* experiment and three additional experiments that are identical to the baseline except in one feature. Below, we describe them in detail. Each name introduced below refers to both the name of an experiment and its treatment. The timeline of an experimental session is explained in the next section.

Baseline (*BL*) involves a substantial amount of misinformation and likely high levels of overconfidence. Thus, we expect it to generate inefficient outcomes based on Propositions 1 and 2. The three main features of *BL* are as follows.

(i) $x = 3$: subject i observes a signal that matches the true state (correct news) if and only if i's quiz score places i in the top one-third of all the subjects in the same experimental session. Thus, two-thirds of the group members receive fake news, and $E(q) = 1/3$ with $\underline{q} < 0.5$.[22]

(ii) *Easy Quiz:* Our theory indicates that a high level of overconfidence makes collective decisions vulnerable to widespread fake news. Prior experimental research suggests that easy tasks are more conducive to overconfidence (i.e., overplacement) than hard tasks.[23] That is, the negative effect of overconfidence on information aggregation is more likely to be borne out with an easy quiz. Our easy quiz is "deceptively simple" since many subjects will obtain relatively high quiz scores, but few will be aware that they are unlikely to receive a correct signal. One relevant question is the following: are policy issues deceptively simple for the average citizen? There is some support for this claim from the political science literature showing that people have strong policy opinions while being somewhat informed, uninformed, or misinformed (see, for example, Kuklinski et al. 2000; Flynn, Nyhan, and Reifler 2017).

(iii) *Informative Voting*: In *BL*, a subject cannot vote for *red* if the subject's signal is *blue*, and vice versa. This is in order to keep the voting game as simple as possible, which reduces behavioral noise and the potential confound of misunderstanding the incentives in the game. In addition, it simplifies the theoretical and experimental analysis of voter behavior.

[22] Recall that if $E(q) \geq 0.5$, overconfidence is unlikely to have an effect on information aggregation in large groups (Section IB).

[23] See, among others, Moore and Cain (2007) and Moore and Healy (2008).

In the remaining three treatments, we relax one feature of *BL* at a time to analyze the effect of (i) a reduction in the share of fake news, (ii) a reduction in the level of overconfidence, and (iii) the ability to vote against signal. These manipulations allow for causal tests of the effect of overconfidence and misinformation. *Top Half* (*TH*) is identical to *BL* except that $x = 2$. That is, subject i observes a signal that matches the true state if and only if i's quiz score is in the top half of all the subjects in the same session. This variation increases the proportion of correct news to one-half (i.e., $E(q) = 1/2$). Both *TH* and *BL* likely involve high levels of overconfidence as assumed in Proposition 2, but only *BL* satisfies its high misinformation condition $E(q) < 1/2$. Thus, a significant increase in the accuracy of group decisions (i.e., efficiency) in *TH* relative to *BL* provides causal evidence in favor of our model. *Harder Quiz* (*HQ*) is identical to *BL* except that its quiz is *not* easy, which will exogenously reduce overconfidence. In other words, both *HQ* and *BL* satisfy the high misinformation condition in Proposition 2, but *HQ* is likely to violate the high overconfidence condition. Thus, a significant increase in efficiency in *HQ* relative to *BL* provides causal evidence in favor of our model. *Strategic Voting* (*SV*) is identical to *BL* except that we remove the informative-voting constraint. Hence, subjects can vote against s_i (e.g., if they are strategic and sufficiently confident that they are in the bottom two-thirds). This treatment allows us to test the effect of informative voting in *BL* as a robustness check.

We now discuss our experimental hypotheses in detail. BNE in *BL* and *HQ* consist of simple cutoff strategies: those who are sufficiently confident that they are in the top one-third should vote for the color of their signal, and the rest should abstain. Note that $\bar{q} \leq 0.5$ is possible in an unbiased electorate.[24] Hypothesis 1 allows for *any* possible q distribution and applies 0.5 as the efficiency benchmark.

HYPOTHESIS 1 (Inefficiency of Overconfidence in *BL* and *HQ*): (*i*) *If the electorate is sufficiently overconfident and rational otherwise*, *the probability of a correct group decision is strictly lower than* 0.5. (*ii*) *Turnout and elicited beliefs regarding the probability of placing in the top one-third are lower*, *and the proportion of correct decisions is higher in HQ than in BL.*

Part (i) of Hypothesis 1 presents a stringent test to infer the negative effect of overconfidence because 0.5 is a distribution-free lower bound for the probability of selecting the correct alternative in a group that consists of unbiased and rational subjects. However, it is also robust to "bounded rationality." For example, the proportion of correct group decisions will still be weakly greater than 0.5 if subjects are unbiased and use a simple heuristic cutoff, such as 50 percent. Thus, Hypothesis 1 is immune to the finding in Esponda and Vespa (2014) that many subjects make suboptimal choices due to not understanding the logic of pivotality.

We predict rampant overconfidence in *BL*, but a harder quiz is likely to stifle overconfidence, which will reduce turnout and improve collective decision-making in *HQ*. Hence, the proportion of correct group decisions is likely lower than 0.5 in

[24] A straightforward example is the case where the quiz is so easy that everyone in the population obtains the same score; i.e., $q_i = \bar{q} = 1/3 < 0.5$ for every i. In that case, every group member must abstain in equilibrium, which results in 0.5 as the probability of a correct group decision.

BL, but this is less likely in *HQ*. The comparison in part (ii) of Hypothesis 1 also reflects this point.

In *SV*, which removes the informative-voting constraint, there is an equilibrium where an unbiased electorate *always* votes for the correct policy due to the experimental structure of signals, regardless of the distribution of q. In this equilibrium, subjects who are *certain* that they are in the top one-third should vote for their signal, and the rest should strategically vote against their signals. Thus, in theory *SV* is starkly different from *BL*. However, in practice such a strategy may be very difficult to coordinate on and unlikely to be followed by subjects. Therefore, Hypothesis 2 applies a less stringent efficiency benchmark based on the following *heuristic* strategy: every i votes for s_i if $p_i(q_i) \geq 0.5$ and against s_i if $p_i(q_i) < 0.5$. In that case, the ex ante probability that an unbiased subject casts a correct vote is higher than 0.66 regardless of the distribution of q (see online Appendix B.9). Thus, our efficiency benchmark for *SV* is 0.66. Given this heuristic strategy, *SV* may improve efficiency unless $p_i(q_i) \geq 0.5$ for too many i with $q_i < 0.5$ (i.e., too many subjects believe that they are more likely to have correct rather than false news).

HYPOTHESIS 2 (Inefficiency of Overconfidence in *SV*): (*i*) *If the electorate is sufficiently overconfident and rational otherwise*, *the probability of a correct group decision is strictly lower than* 0.66. (*ii*) *The proportion of correct decisions is higher in SV than in BL* (*unless the electorate is extremely overconfident*).

TH is identical to *BL* in every aspect except that $x = 2$ (i.e., s_i matches the true state if and only if i is in the top half of all the subjects in the same session). Once again, BNE consists of cutoff strategies: subjects should vote for s_i if they are sufficiently confident that they are in the top half and abstain otherwise. In *TH*, the probability of a correct group decision is strictly greater than 0.5 in groups that are rational and unbiased.[25] Part (i) of Hypothesis 3 reflects this point. Part (ii) follows since *TH* does not satisfy the condition $E(q) < 1/2$ in Proposition 2, unlike *BL*.

HYPOTHESIS 3 (Inefficiency of Overconfidence in *TH*): (*i*) *If the electorate is sufficiently overconfident and rational otherwise*, *the probability of a correct group decision is weakly lower than* 0.5. (*ii*) *The proportion of correct decisions is higher in TH than in BL.*

Hypotheses 1–3 concern the efficiency of voting outcomes in our treatments, allowing for overconfident beliefs but assuming that subjects are otherwise (boundedly) rational. Hypotheses 4 and 5 below provide tests for the rationality of turnout behavior. In particular, we check whether the probability of voting for s_i increases in $p_i(q_i)$, consistent with a cutoff strategy. We specifically focus on 50 percent as an intuitive cutoff for turnout.

[25] In *TH*, $\bar{q} < 0.5$ cannot hold because unbiased beliefs must average out to 0.5. Thus, $\bar{q} \geq 0.5$. However, $\bar{q} = 0.5$ if and only if $q_i = \bar{q} = 0.5$ for every i. Ruling out this knife-edge case, $\bar{q} > 0.5$, and thus the probability of a correct group decision is strictly greater than 0.5.

HYPOTHESIS 4 (Rationality of Turnout in Treatments): (*i*) *Subjects who report higher* $p_i(q_i)$ *are more likely to vote.*[26] (*ii*) *Subjects who vote report* $p_i(q_i) \geq 0.5$.

The voting part in each experiment starts with a control condition, which allows us to check whether subjects understand (or learn) the incentives in the voting game. This condition, called *Objective Information* (*OBJ*), mirrors the treatments, but s_i and q_i are objective information rather than based on subjective beliefs. In every round, q_i for subject i is a new i.i.d. draw from the uniform distribution between 0 and 1, which is common knowledge. Prior to the decision whether or not to vote, subject i privately observes q_i and then s_i, which is either *red* or *blue* as in the treatment. Hence, we can study how turnout depends on q_i. *OBJ* is identical in *BL*, *HQ*, and *TH*, and imposes informative voting, whereas voting against s_i is allowed in the *OBJ* condition of *SV*.

HYPOTHESIS 5 (Rationality of Turnout in *OBJ*): (*i*) *Subjects with higher* q_i *are more likely to vote.* (*ii*) *Subjects who vote have* $q_i \geq 0.5$.

B. *Experimental Protocol*

As shown in Figure 1, every session begins with a phase to familiarize subjects with the quadratic scoring rule, which we use to elicit beliefs after the quiz and during the voting part.[27] Next, subjects take the quiz on math and logic puzzles. In this phase, subjects do not know about the voting part or the elicitation of beliefs regarding their quiz performance. The quiz consists of 20 questions and has a time limit of 10 minutes. Subjects are informed that each correct answer is rewarded by €0.30. At the end of the quiz, subjects are asked to report the probability that their quiz score places them in the top $1/x$ of all the subjects in the same session, where $x = 3$ in *BL*, *HQ*, and *SV*, and $x = 2$ in *TH*. Subjects earn up to €3 in this task.[28] This is followed by the voting part, which starts with a control condition, i.e., *OBJ*. The aim of *OBJ* is to teach the voting game using an objective information structure and to test Hypothesis 5. After every round, subjects receive detailed feedback about the round's outcome including the accuracy of the group decision and whether or not their signal was correct.

OBJ is followed by the treatment, in which subjects know that their performance in the quiz (that is, whether or not they are in the top $1/x$) determines their signal. Thus, each i makes choices based on a subjective belief $p_i(q_i)$, which we elicit after the quiz as discussed above. Before the treatment begins, subjects are reminded of their $p_i(q_i)$ report at the end of the quiz. Each treatment has six rounds. Subjects know that from the first five rounds, one round is randomly selected for payment, and that the final round is a high-stake round as earnings in that round are always paid out. Each group member receives €5 from a correct group decision in paid

[26] In *SV*, voting refers to voting for s_i as subjects in *SV* can vote against s_i.

[27] One caveat is that elicited beliefs may tend towards the center if the mechanism is quadratic scoring rule or its successor, binarized scoring rule (see, e.g., Danz, Vesterlund, and Wilson forthcoming).

[28] It is explained to subjects that out of N individuals in the session exactly N/x are in the top $1/x$; i.e., they know that if there are score ties, they are broken randomly (N/x is always an integer number). Our data analysis uses this final placement of subjects in the top $1/x$ after ties (if any) are broken, as implemented during the experiment.

Phase 1: Teaching belief elicitation	Phase 2: Measures of competence and beliefs	Phase 3: Voting with objective information (control condition)	Phase 4: Voting with subjective information (treatment)
• Instructions on quadratic scoring rule • Belief elicitation about several true or false statements (e.g., trivia)	• Incentivized quiz with 20 math and logic questions: measure of competence q in treatment • Belief elicitation: measure of belief $p(q)$ in treatment	• q: i.i.d. draw from $U[0, 1]$ for every subject in every round • Belief elicitation about other members' propensity to vote • Detailed feedback after every round	• q depends on subject's quiz performance in Phase 2 (in particular, i's signal is correct if and only if i is in the top $1/x$) • Subjects reminded about their elicited belief in Phase 2, $(p(q))$ • Belief elicitation about other members' propensity to vote • Final round: belief elicitation about (other) voters' q • No feedback (except in 6 *BL* sessions with aggregated feedback)

Figure 1. Timeline of Experimental Sessions

Notes: Harder Quiz (*HQ*) differs from *Baseline* (*BL*) in Phase 2. *Top Half* (*TH*) differs from *BL* in Phases 2 and 4. *Strategic Voting* (*SV*) differs from *BL* in Phase 4 (also, Phase 3 in *SV* allows for strategic voting).

rounds. There are two belief elicitation tasks about *other* group members in each treatment. These beliefs are relevant for a subject's theoretical best response and concern (i) the probability that a randomly selected other group member chooses to vote, and (ii) the probability that a randomly selected (other) voter is in the top $1/x$ of all the subjects in the same session, where $x \in \{2,3\}$ depending on the treatment (in *SV*, we ask the probability that a randomly selected voter votes correctly).[29] Subjects do not receive feedback after making their choices in the treatments, except in six *BL* sessions in which we provide subjects with aggregated feedback after each round regarding the accuracy of the group decision and the total number of voters, but not the true state since this would reveal subjects' placement in the quiz.[30]

We ran 12 *BL* sessions with $N = 24$ before the COVID-19 pandemic. After the pandemic began, we lowered N to 15 in *BL* as the lab capacity was reduced. In particular, $N = 16$ in *TH*, and $N = 15$ in *HQ*, *SV*, and eight *BL* sessions.[31] We group these 8 *BL* sessions into *Baseline* 15 (*BL* 15) and pool its data with 12 *Baseline* sessions with $N = 24$—unless otherwise stated—since main results in sessions

[29] We elicited beliefs about other voters' quiz performance only in the last round so as not to overburden subjects in every round with two belief elicitation tasks. Subjects are asked to report the probability that a randomly selected group member chooses to vote also in *OBJ*.

[30] More generally, there are three design variations within the *BL* sessions. These variations are minor in the sense that they give rise to robust main results as shown in online Appendix B.1. The first variation gives subjects aggregated feedback in six sessions as described above. The second variation reduces N from 24 to 15 as explained in more detail below. The third variation slightly increases quiz difficulty in two sessions with $N = 15$.

[31] Due to a coding error in one quiz answer in earlier *HQ* sessions, one subject was wrongly ranked in the top one-third, and another was wrongly ranked in the bottom two-thirds in one *HQ* session. The error cannot have influenced subjects' behavior as there is no feedback in any part of the session that concerns the quiz or placement in the top one-third. Therefore, we manually corrected the implications of this error in the data.

TABLE 1—EXPERIMENTAL DESIGN

Experiment/treatment	Easy quiz	Top $1/x$	Voting against s_i	$P_{correct}$	N	No. sessions
Baseline (*BL*)	Yes	1/3	No	≥ 0.5	24 or 15	20
Harder Quiz (*HQ*)	**No**	1/3	No	≥ 0.5	15	6
Strategic Voting (*SV*)	Yes	1/3	**Yes**	$\geq 2/3$	15	6
Top Half (*TH*)	Yes	**1/2**	No	> 0.5	16	6

Notes: This table shows the main features of treatments. Major design differences are indicated in bold. The column $P_{correct}$ concerns the "efficiency benchmark" and refers to the theoretical probability that a group of (boundedly) rational and unbiased individuals makes the correct decision. Fake news share increases in x as explained in the text. In *TH*, $N = 16$ so that an integer number of subjects is in the top one-half. *BL* sessions involve small variations (see footnote 29 and online Appendix B.1). All treatments have a control condition with objective information (*OBJ*).

with $N = 15$ and $N = 24$ are not statistically different (see online Appendix B.1). *Baseline* (*BL*) refers to the pooled data of these 20 sessions.[32]

Table 1 summarizes the main features of our experiments. In total, we have 38 sessions with 684 subjects. Sessions are conducted at the Vienna Center for Experimental Economics using z-Tree (Fischbacher 2007). Subjects are recruited from the general undergraduate population, and recruitment is gender balanced. In both the control condition and treatment, subjects have to correctly answer a set of control questions in order to proceed to the game. The average payoff per subject is €20.90.[33]

III. Experimental Results

The presentation of experimental findings is organized as follows. We begin with a discussion of aggregate results on turnout, elicited beliefs, and the accuracy of group decisions to address Hypotheses 1–3. Section IIIB investigates voter overconfidence. In Section IIIC, we analyze individual turnout behavior and its determinants in detail.

Statistical tests are based on nonparametric statistics and conducted at the session (i.e., electorate) level. We report Mann-Whitney p-values for pairwise treatment comparisons. When conducting paired-data and one-sample tests, we use both the sign test and the Wilcoxon signed-rank test. In the main text, we report only the sign test p-values unless the two tests disagree (see online Appendix B.3 for the test results that are not reported in the text). We also use parametric analysis where appropriate. For example, we use random effects probit estimations for the analysis of individual turnout behavior in Section IIIC.

[32] We also had to have shorter sessions due to pandemic restrictions, and we reduced the number of rounds in the *OBJ* condition from 15 in the prepandemic sessions to 8 in the postpandemic sessions. We pool the data of *OBJ* in all sessions as the main results in *OBJ* do not statistically differ due to the difference in the number of rounds and N (see online Appendix B.2).

[33] See instructions in online Appendix C.

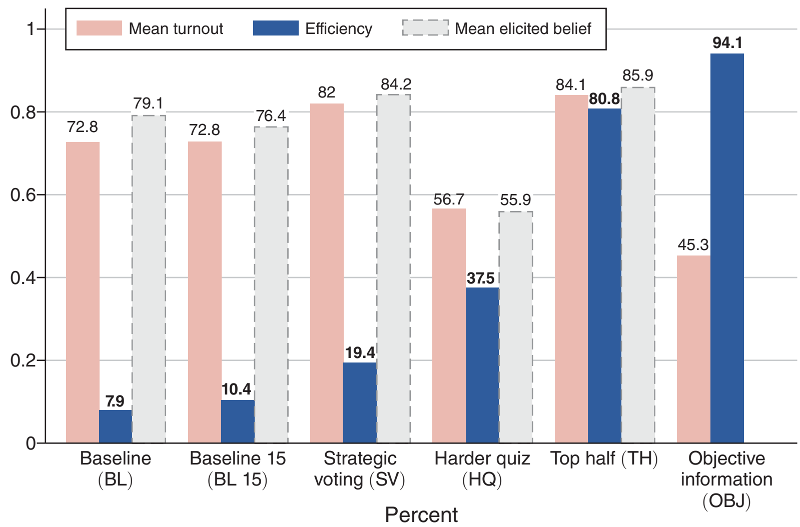

Figure 2. Efficiency, Elicited Beliefs, and Turnout by Treatment

Notes: The share of subjects who receive correct news is one-third in *BL*, *BL* 15, *SV*, and *HQ*, and one-half in *TH*. "Mean elicited belief" is the average elicited belief about the likelihood of scoring in the top $1/x$ (i.e., the average subjective probability of receiving correct news). Turnout data for *SV* (and *OBJ* in *SV*) considers only those who vote for s_i.

A. *Efficiency*, *Turnout*, *and Elicited Beliefs*

Figure 2 shows the average proportion of correct group decisions (i.e., efficiency), the average elicited belief regarding the likelihood of scoring in the top $1/x$ (i.e., the subjective probability of receiving a correct signal), and the average turnout rate by treatment. For transparency, we distinguish between *Baseline* (*BL*) and *Baseline* 15 (*BL* 15). Figure 2 also presents the efficiency and turnout rates in the pooled data of *OBJ*. Throughout Section IIIA, we test directional hypotheses and report one-sided p-values.

We observe a drastic inefficiency in *BL* and *BL* 15 consistent with Hypothesis 1.

RESULT 1 (Inefficiency is rampant in *BL* and *BL* 15): *The percentage of correct group decisions is 7.9 percent in BL and 10.4 percent in BL 15, which is lower than the 50 percent threshold* ($p < 0.001$ *and* $p = 0.004$).[34]

We explore the reason for this drastic failure of information aggregation by analyzing turnout and elicited beliefs. Figure 2 shows that the turnout rate is 72.8 percent in both *BL* and *BL* 15, which is excessively high because a turnout rate above two-thirds in a round is guaranteed to result in the wrong choice when only one-third of group members receive correct news. In turn, the excessive turnout is driven

[34] As mentioned above, we use the Mann-Whitney test for pairwise treatment comparisons, and we use both the sign test and the Wilcoxon signed-rank test in one-sample or paired-data tests. We report the sign test result unless the two tests disagree. See online Appendix B.3 for the test results that are not reported in the text.

by overconfident beliefs, as we explain in Section IIIC. Indeed, one striking feature of Figure 2 is the close comovement of average turnout and elicited beliefs across treatments. This point will also become clear below as we discuss Result 2.

RESULT 2 (*HQ* attenuates overconfidence and inefficiency): *The percentage of correct group decisions is 37.5 percent in HQ, which is higher than that in BL and BL 15* ($p = 0.004$ *and* $p = 0.023$, *respectively*). *HQ reduces turnout and mean* $p_i(q_i)$ *as predicted. We find no evidence that efficiency in HQ is lower than 50 percent* ($p = 0.109$).

HQ replicates *BL* with a harder quiz to reduce overconfidence and test whether this reduction improves collective decision-making. We find that *HQ* reduced the average quiz score by about 5.[35] As a result, mean $p_i(q_i)$ decreased to 55.9 percent in *HQ* from 76.4 percent in *BL* 15 ($p = 0.005$), and 79.1 percent in *BL* ($p < 0.001$), as Figure 2 shows. Among subjects placed in the bottom two-thirds, mean $p_i(q_i)$ fell by more than a third.[36]

The decline in the elicited beliefs in *HQ* is associated with a decrease in turnout. As Figure 2 shows, the average turnout rate fell from 72.8 percent (in both *BL* and *BL* 15) to 56.7 percent in *HQ* ($p = 0.008$ and $p = 0.016$, respectively). Also, *HQ* significantly reduced the "belief gap among voters," which is defined and discussed in detail in Section IIIB below.

As a result of these effects on beliefs and turnout, efficiency more than tripled in *HQ*, which shows that reducing overconfidence causally improves the quality of democratic choice. We note that there is substantial variation in the proportion of correct group decisions across sessions of *HQ* (ranging from 0 to 92 percent). This variation in efficiency is nicely explained by the sizable differences in overconfidence at the session level. To show this point, we use a regression analysis. We describe the analysis and report its results in online Appendix B.5.

Next, we consider the effect of *SV* on efficiency. While *SV* did increase efficiency to some extent, this difference is not statistically significant. Thus, our baseline results are robust to allowing for strategic voting.

RESULT 3 (*SV* does not mitigate inefficiency): *The percentage of correct group decisions is 19.4 percent in SV, which is not significantly different from that in BL or BL 15* ($p = 0.133$ *and* $p = 0.315$, *respectively*). *In addition, it is lower than the 66 percent threshold* ($p = 0.016$).

The lack of evidence for increased efficiency under *SV* is despite all subjects with $p_i(q_i) < 0.5$ (strategically) voting against s_i and due to substantial overconfidence (i.e., few subjects report that $p_i(q_i) < 0.5$). Also, we observe a notable variation in the proportion of correct group decisions across sessions of *SV* (ranging from 0 to 58.9 percent). Similar to our observation for *HQ*, this variation is well explained by the differences in overconfidence across *SV* sessions. To show this point, we

[35] More precisely, the quiz score decreased to 11.18 in *HQ* from 16.16 in *BL* 15 and 16.54 in *BL* ($p < 0.001$ in both). The distribution of quiz scores can be found in online Appendix B.4.

[36] It is 45.4 percent in *HQ* versus 68.6 percent in *BL* 15 ($p = 0.007$) and 72.4 percent in *BL* ($p < 0.001$).

use the data of *SV* and the regression analysis mentioned above for *HQ* (see online Appendix B.5).

TH increases the share of correct news from one-third to one-half. We find that this increase substantially improves efficiency, which is in line with our theoretical framework since the condition $E(q) < 1/2$ in Proposition 2 is not satisfied.

RESULT 4 (Reduced fake news in *TH* increases efficiency): *The percentage of correct group decisions is 80.8 percent in TH, which is significantly higher than that in BL and BL 15* ($p < 0.001$ *in both cases*). *There is evidence that efficiency in TH exceeds 50 percent according to the Wilcoxon signed-rank test* ($p = 0.022$), *but not the sign test* ($p < 0.109$).[37]

Result 4 points to the importance of limiting the share of misinformation for information aggregation. In particular, it provides causal evidence that reducing the extent of fake news in circulation improves the quality of democratic choice in our framework. Nevertheless, we also note that elections in *TH* tend to be very close; i.e., the correct policy is typically chosen with a small margin. If we had a sizable partisan minority in the design of *TH* as assumed in Proposition 3, there could be many more wrong decisions. More generally, our theoretical extensions in Section IC are relevant for showing how overconfidence can undermine information aggregation even when $E(q) \geq 1/2$, and they can be implemented and tested in future experimental work.

B. *Voter Overconfidence*

In Section IIIA, we have shown that increasing quiz difficulty reduces overconfidence and increases efficiency (*BL* versus *HQ*). Furthermore, 95.3 percent of voters across all treatments report that $p_i(q_i) \geq 0.5$. That is, too many subjects vote believing that they are more likely to be correct than wrong. These observations provide suggestive evidence for overconfidence among voters. To investigate the extent and prevalence of voter overconfidence formally, we use the following test based on the work by Benoît and Dubra (2011) and Benoît, Dubra, and Moore (2015). We first define the *belief gap among voters* as the difference between the mean $p_i(q_i)$ among voters in a session and the actual share of voters in that session who score in the top $1/x$ (i.e., we disregard "abstainers"). The belief gap among voters can be positive or negative. A statistically significant positive belief gap among voters is compelling evidence for voter overconfidence.[38]

Table 2 shows that the belief gap among voters is statistically significant in all four treatments.[39] The belief gap among voters is substantial in *BL* and *SV* consistent

[37] The sign test result is due to one *TH* session with high overconfidence coupled with abstentions from subjects who place in the top half, reducing efficiency in that session below 0.5.

[38] As mentioned above, we use both the sign test and the Wilcoxon signed-rank test in paired-data and one-sample tests. Throughout this section, we conduct two-sided tests. We report only the sign test result unless the two tests disagree. See online Appendix B.3 for the test results not reported in the main text.

[39] Some experimental studies have found that men tend to be more overconfident than women. The belief gap does not differ across male and female voters in our treatments ($p = 0.115$ in *BL* and $p = 0.688$ in *HQ*, *SV*, and *TH*).

Table 2—Voters' Elicited Beliefs versus Actual Placement in the Top $1/x$

Treatment	Top $1/x$	Mean $p_i(q_i)$ among voters	Share of voters in the top $1/x$	Belief gap among voters (p-value)
Baseline (*BL*)	1/3	86.5%	40.0%	46.4% ($p < 0.001$)
Harder Quiz (*HQ*)	1/3	68.9%	46.0%	22.9% ($p = 0.031$)
Strategic Voting (*SV*)	1/3	89.5%	36.3%	53.2% ($p = 0.031$)
Top Half (*TH*)	1/2	90.0%	54.8%	35.2% ($p = 0.031$)

Notes: The belief gap among voters equals the difference between the second column and the third column (see the main text for details). The share of voters in the top $1/x$ may differ from $1/x$ because abstainers are excluded.

with their high inefficiency.[40] The belief gap in *HQ* is significantly different from 0 but much smaller than that in *BL* ($p < 0.001$) or *SV* ($p = 0.004$) due to the difficulty of the quiz.

We note that subjects in the top one-third in *BL*, *HQ*, and *SV* jointly state higher beliefs (88.2 percent) than those who are not (69.5 percent), which is significant ($p < 0.001$). The analogous result holds in *TH*: 95.1 percent versus 76.7 percent ($p = 0.031$). Thus, elicited beliefs are informative in the sense that placement in the top $1/x$ is associated with higher elicited beliefs.

In the last round of each treatment, we ask subjects to report the probability that a randomly selected (other) voter is in the top $1/x$, which we denote by "$p_i(q_j \mid vote_j = 1)$."[41] In the main text of the theory section, we made the simplifying assumption that individuals are unaware of other individuals' overconfidence; that is, they are overconfident in the competence of the average voter in equilibrium. We find that subjects indeed exhibit such overconfidence, justifying our assumption. In treatments with $x = 3$ and no aggregated feedback, mean $p_i(q_j \mid vote_j = 1)$ is 54.7 percent, whereas the average share of competent voters in the last round is 40.3 percent. This difference is statistically significant ($p < 0.001$) but also moderate relative to subjects' and voters' own overconfidence. We consider six *BL* sessions with *aggregated feedback* separately as subjects learn the accuracy of the group decision and turnout level at the end of every round. Such feedback seems to make subjects better calibrated regarding the competence of the average voter. Mean $p_i(q_j \mid vote_j = 1)$ is 47.9 percent with 41.5 percent of voters in the top one-third in the last round ($p = 0.219$ according to the sign test; $p = 0.075$ according to the Wilcoxon signed-rank test).[42]

C. *Determinants of Individual Turnout and Mistakes*

We now present a more detailed analysis of the relationship between turnout and elicited beliefs. Figure 3 shows the turnout rate for several intervals of beliefs as well as for $p_i(q_i) = 0$ because elicited beliefs are highly concentrated

[40] The belief gap is slightly inflated in *SV* because we consider only subjects who vote for s_i. About 11.4 percent of voters in *SV* vote against s_i (in particular, all subjects with $p_i(q_i) < 0.5$ vote against s_i).

[41] As mentioned in Section IIB, we ask subjects in *SV* to report the probability that a randomly selected (other) voter votes correctly because subjects can vote against their signals.

[42] In one of the *TH* sessions, the treatment terminated at the end of the fifth round due to a mistake. Therefore, only five *TH* sessions provide data for the variable $p_i(q_j \mid vote_j = 1)$. Mean $p_i(q_j \mid vote_j = 1)$ is 70.2 percent, whereas the average share of voters in the top half is 54.6 percent ($p = 0.063$ according to the sign test; $p = 0.043$ according to the Wilcoxon signed-rank test).

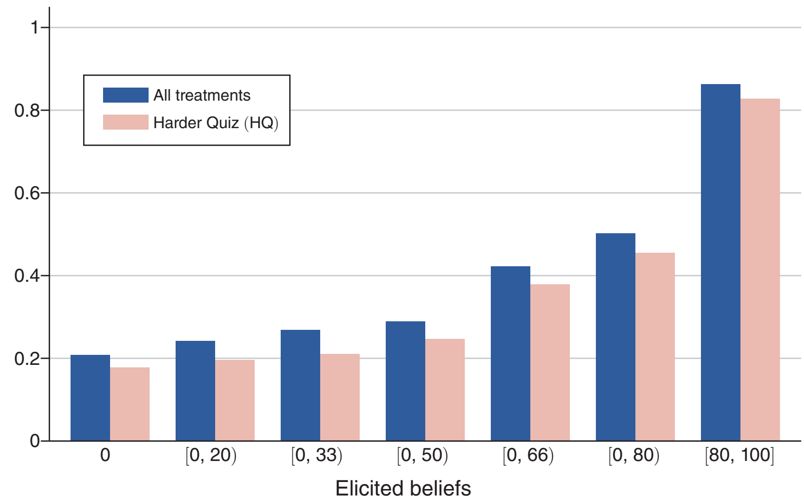

Figure 3. The Relationship between Treatment Turnout Rate and Elicited Beliefs

Notes: The figure shows the turnout rate for the given value or intervals of elicited beliefs. Turnout data in *Strategic Voting* includes only those who vote for their signal.

on certain values. For example, among subjects with $p_i(q_i) < 0.5$, $p_i(q_i) = 0$ for about 44 percent, and close to 75 percent report that $p_i(q_i) \in \{0, 0.3, 0.4\}$. We single out *HQ* in Figure 3 because it has a significantly larger share of subjects with $p_i(q_i) < 0.5$ in comparison to other treatments due to its harder quiz.

The monotone pattern in Figure 3 shows that turnout depends on elicited beliefs. Figure 7 in online Appendix B.6 replicates this pattern and suggests that 50 percent is empirically relevant for voting and abstention choices as a heuristic cutoff. To investigate the determinants of voting behavior and address Hypothesis 4 formally, we estimate a random effects panel probit model. The dependent variable is the individual decision whether or not to vote in the treatment. The independent variables in specifications (1) and (2) are (i) the elicited belief regarding the subject's own placement in the top $1/x$ ("$p_i(q_i)$"), (ii) the elicited belief regarding other group members' likelihood of voting ("$p_i(vote_j = 1)$"), (iii) placement in the top $1/x$ ("top $1/x$"), and (iv) a time trend. We cluster errors at the session level. Table 3 reports the average marginal effects.

Comparing specifications (1) and (2) in Table 3, we note that including dummy variables for treatment variations as "controls" makes no difference.[43] The effect of the variable "$p_i(q_i)$" is positive and highly significant ($p < 0.001$), consistent with theory. The coefficients in (1) and (2) show that on average, a 1 percent increase in $p_i(q_i)$ increases the probability of voting by 0.5 percent. The effect of the variable "$p_i(vote_j = 1)$" is positive and highly significant ($p < 0.001$). This suggests that

[43] Controls are dummy variables that represent *SV*, *HQ*, *TH*, and the baseline variation with *aggregated feedback*. Aggregated feedback has an effect on the variable "$p_i(q_j | vote_j = 1)$," as discussed in Section IIIB, and on turnout, as discussed in online Appendix B.1.

Table 3—Explaining Individual Turnout Decision

	All rounds (1)	All rounds (2)	Final round (3)	Final round (4)
$p_i(q_i)$	0.492	0.487	0.487	0.462
	(0.031)	(0.036)	(0.045)	(0.050)
$p_i(vote_j)$	0.251	0.244	0.349	0.339
	(0.038)	(0.035)	(0.077)	(0.076)
top $1/x$	0.097	0.093	0.095	0.092
	(0.024)	(0.024)	(0.039)	(0.038)
$p_i(q_j \| vote_j = 1)$	—	—	0.148	0.126
			(0.064)	(0.065)
Round number	−0.012	−0.012	—	—
	(0.003)	(0.003)		
Controls	No	Yes	No	Yes
Observations	4,088	4,088	668	668

Notes: Coefficients show the average marginal effects in random effects panel probit regressions described in the main text. The dependent variable is the subject's binary choice between voting (= 1) and abstaining/voting against s_i in *SV* (= 0). Standard errors clustered by session are in parentheses.

expecting higher turnout from other group members encourages turnout.[44] The effect of the variable "top $1/x$" is also positive and highly significant, likely because it captures variation in beliefs not fully accounted for by subjects' $p_i(q_i)$ reports as well as (numerical) errors subjects may make in entering $p_i(q_i)$.[45]

Columns 3 and 4 report results from probit regressions using only the data of the final round with the variable "$p_i(q_j | vote_j = 1)$" as an additional regressor. The final round is a high stake round since subjects earn €5 from a correct group decision in this round (see Section IIB). The results are consistent with the results of (1) and (2). Including dummy variables for treatment variations makes no difference except for the variable "$p_i(q_j | vote_j = 1)$," the effect of which becomes marginally significant with controls.[46]

As mentioned above, across all treatments 95.3 percent of voters report that $p_i(q_i) \geq 0.5$. Similarly, voting is very rare in *OBJ* once q_i falls below 50 percent (only 9.9 percent of individuals with $q_i < 0.5$ vote). For this reason, the average turnout rate in *OBJ* remains at a low 45.3 percent, and the average efficiency is very high at 94.1 percent, as Figure 2 shows. Moreover, our random effects panel probit

[44] A theoretical mechanism to explain this result is as follows. Higher turnout among other members is associated with a lower cutoff strategy, which implies reduced expected accuracy among other voters. Although higher turnout is potentially beneficial (i.e., if the expected accuracy among other voters is above 0.5), the reduction in the expected accuracy may dominate and reduce the subject's best-response cutoff.

[45] As shown in Figure 3 and Figure 7 in online Appendix B.6, even subjects with high $p_i(q_i)$ abstain. A comparison of the two panels of Figure 7 indicates that this pattern is more frequent among subjects who rank below the top $1/x$, suggesting that subjects in the top $1/x$ with high $p_i(q_i)$ are firmer in their confidence, less likely to have "second thoughts" regarding their ranking, and thus less likely to abstain.

[46] One possible interpretation is as follows. Voters tend to be highly overconfident, and it may be natural for overconfident voters to also overestimate the competence of other voters, giving rise to the positive and statistically significant effect of "$p_i(q_j | vote_j = 1)$" in (3). Another possibility is that this effect is related to the six *BL* sessions with aggregated feedback, which reduced $p_i(q_j | vote_j = 1)$ as explained in Section IIIB. Indeed, the controls in specification (4) in Table 3 make the effect of $p_i(q_j | vote_j = 1)$ marginally significant. Also, excluding the data of the six sessions with feedback in the regression makes its coefficient insignificant.

model estimating the propensity to vote in *OBJ* shows that it strongly increases in q_i (see online Appendix B.2). Combining all of these findings, we obtain supporting evidence for Hypotheses 4 and 5.

RESULT 5: (*i*) *Most subjects who vote in the treatments report* $p_i(q_i) \geq 0.5$, *and the propensity to vote increases in* $p_i(q_i)$ *and when placed in the top* $1/x$. (*ii*) *Most subjects who vote in OBJ have* $q_i \geq 0.5$, *and the propensity to vote increases in* q_i.

We now focus on the behavior of subjects with low confidence (i.e., subjects who report $p_i(q_i) < 0.5$) in the treatments. This is a small group of 82 subjects (out of a total of 684). Voting despite $p_i(q_i) < 0.5$ can be loosely interpreted as a mistake.[47] Figure 3 shows that even some subjects who indicate that they are *certain* to be below the top $1/x$ vote possibly due to errors in understanding the game or in entering beliefs or voting choices (20.8 percent overall and 17.8 percent in *HQ*). However, a large fraction of subjects with $p_i(q_i) < 0.5$ make relatively few mistakes. In particular, two-thirds $(55/82)$ of subjects with low confidence vote at most twice with an average turnout rate of 10 percent.[48] The remaining subjects frequently make mistakes not only in the treatment but also in *OBJ*. To formalize this point, we estimate a random effects panel probit model in which the dependent variable is the individual voting decisions of these 82 subjects in the treatments. The new explanatory variable is the "mistake rate in *OBJ*," defined as the turnout rate conditional on $q_i < 0.5$. As before, we control for quiz performance and elicited beliefs, and in some regression specifications, we use dummies to control for treatment variations (see online Appendix B.7). In every specification, we find that the coefficient of the mistake rate in *OBJ* is statistically and economically significant $(p < 0.001)$. Depending on the regression specification, a 1 percent increase in the mistake rate in *OBJ* is associated with a turnout rate increase of 0.6–1.2 percent among subjects with $q_i < 0.5$ in the treatment. This evidence suggests that understanding *OBJ* implies understanding the incentives in the treatment. This is reassuring for our claim that subjects are largely rational in the treatments as a substantial majority of subjects make no or few mistakes in *OBJ*.

D. *Robustness*

We have not adjusted the *p*-values for multiple comparisons in our analysis as we are testing a formal theoretical framework. A convincing statistical statement in support of a theoretical framework requires (almost) all experimental results to be consistent with it, which is an inherent statistical correction mechanism. In other words, obtaining some empirical results consistent with the theory (along with several or many inconsistent results) does not constitute convincing support.[49] Still, we

[47] The caveat is that if the wrong alternative is more likely to lead in votes, voting for s_i given $p_i(q_i) < 0.5$ is not a definite mistake in theory.

[48] We also note that four subjects who report $p_i(q_i) < 0.5$ are actually placed in the top $1/x$ and have an average turnout rate of 37.5 percent.

[49] Among the various multiple-testing environments, theory testing is more similar to "intersection-union testing" (Berger 1982) also called "conjunction testing" by Rubin (2021), which does not require a correction for multiple testing. In intersection-union testing, the null hypothesis is the union of two or more component hypotheses, and the null is rejected only when all the component hypotheses are rejected.

can show the robustness of our pairwise treatment comparisons to well-known multiple testing correction procedures. Since we have significantly lower power with *BL* 15, our preferred baseline group in this exercise is *BL*. Our experimental hypotheses consider the comparison of multiple treatments (*BL*, *HQ*, *SV*, and *TH*) and multiple outcomes in the comparison of *BL* and *HQ*, since we hypothesize that *HQ* improves outcomes due to lower quiz scores, lower mean $p_i(q_i)$, and reduced turnout. Table 9 in online Appendix B.8 presents the adjustment of the *p*-values of these six pairwise comparisons using Bonferroni, Holm, and Benjamini and Hochberg methods. Our main conclusions remain unchanged in each method (see online Appendix B.8).

IV. Related Literature and Concluding Remarks

In this paper, we delineate conditions under which the joint effect of overconfidence and misinformation harms information aggregation assuming that individuals make otherwise rational choices. These conditions would typically be relaxed under bounded rationality, such as cursed voting strategies (Eyster and Rabin 2005) and heuristic cutoff strategies.

Information Aggregation.—In our setup, individuals differ in their competence and may choose to abstain to improve information aggregation.[50] Therefore, our model is closely related to Feddersen and Pesendorfer (1996) and its various theoretical extensions, such as Feddersen and Pesendorfer (1999); McMurray (2013); and Herrera, Llorente-Saguer, and McMurray (2019a); as well as several experimental studies, which analyze among other things whether less-informed or uninformed individuals strategically abstain (see, e.g., Battaglini, Morton, and Palfrey 2008, 2010; Morton and Tyran 2011; Bhattacharya, Duffy, and Kim 2014; Mengel and Rivas 2017; Herrera, Llorente-Saguer, and McMurray 2019b; and Elbittar et al. 2020). Our study differs from this literature in that we focus on the effect of overconfidence and misinformation on voting behavior and information aggregation.[51]

Overconfidence.—There is an extensive literature analyzing the impact of overconfidence on individual decision-making as well as on markets.[52] Few studies focus on the role of cognitive biases, such as overconfidence, in voting and collective decision-making. Important exceptions are Levy and Razin (2015) and Ortoleva and Snowberg (2015). In these models, each individual receives repeated information signals regarding the state of the world, which may be correlated. Unlike in our model, individuals do not differ in their competence, but in the degree to which they underestimate the correlation between their signals. Ortoleva and Snowberg (2015) show, theoretically and empirically, that overconfidence arising from correlation

[50] Our results also apply to committee decision-making. Committees often consist of professionals, who may be overconfident regarding their expertise. According to Kahneman (2011), "overconfident 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."

[51] The experiment by Morton, Piovesan, and Tyran (2019) also studies voting with misleading information and overconfidence. However, their study has a different focus, does not allow for abstention, and does not provide a theoretical analysis of information aggregation with overconfidence.

[52] See, among others, Camerer and Lovallo (1999); Malmendier and Tate (2005); DellaVigna (2009), and references therein.

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 mitigating the impact of partisan preferences on voting. This seemingly contrasts with our results and in particular Proposition 3. However, there are several notable differences between Levy and Razin (2015) and our study. The source of the cognitive bias is different, individuals are heterogeneous in competence, and abstention is allowed in the current study. Levy and Razin (2015) implement a nuanced form of partisan preferences, whereas our Proposition 3 involves only nonpartisans and full partisans. Therefore, our findings complement theirs.[53]

Misinformation.—Our paper focuses on the cognitive mechanism behind the influence of misinformation on judgments and opinion formation. An individual's judgment may also be affected by their ideology and partisanship even when they desire to have an accurate opinion (Kunda 1990). Hence, there are two possible mechanisms underlying the influence of misinformation on judgments: a motivated-reasoning mechanism suggesting that the effect of misinformation depends on ideology or partisanship, and a cognitive mechanism suggesting that it depends on analytic reasoning and critical thinking skills, which we denote by "competence" in this paper.[54] While our focus is on the role of competence and the related overconfidence, our framework may be extended to relate to partisanship. Even in the absence of a partisan motive, overconfidence and misinformation may generate partisan-like opinions and behavior that hardly vary across the two states of the world, as Example 1 illustrates. Thus, in a dynamic context, the emergence of partisanship and the related directional goals could be—at least in part—explained by cognitive factors and overconfidence.[55]

Concluding Remarks.—The ideas put forward in this paper are relevant for various forms of civic and political participation. Not only voting, but also protest participation, petitions, costly information acquisition, communication, and opinion leadership can serve as mechanisms of information aggregation and are likely subject to similar misinformation and overconfidence problems as in our framework.

[53] In contrast to our inefficiency results, a social learning literature shows that overconfidence may promote social learning by breaking incorrect information cascades or by making cascades emerge later (see, e.g., Bernardo and Welch 2001; and Goeree et al. 2007). An important difference to our model is that this literature does not consider the possibility of $\underline{q} < 0.5$ or third-party sponsored news, which are relevant scenarios due to reasons discussed in Sections IB and IC as well as the global emergence of nonprofessionals and nonexperts, such as celebrities, athletes, and social media influencers as opinion leaders on a wide range of topics. In addition, the experimental literature has not implemented subjective beliefs in the way we do, but doing so may be a fruitful direction for future research.

[54] Our model can be extended in future work to allow for heterogeneous (i.e., motivated) priors regarding the state of the world as well as motivated updating from signals. We conjecture that similar results to those in this paper will hold under the assumption that a specific prior is correct based on the best available information at the time opinions are being formed.

[55] Closely related in this respect is Ortoleva and Snowberg (2015), who show that overconfidence arising from correlation neglect can lead to stronger partisanship as discussed above.

Analyzing a richer model with a larger scope of political participation is an important direction for future research.

We note that our results are of no support for scholarly and popular arguments against democracy, which have increased in prominence in the last decade. Instead, our results point to the extremely important role of institutions in upholding democratic principles and processes. These institutions, especially high-quality education, must be accessible and in the service of the public. In many countries, a succession of political and corporate scandals, financial crises, and ever-widening inequalities have resulted in growing distrust and the perception that the current institutions serve a select group rather than the public at large. Such distrust renders the propagation of fake news easier, overconfidence more dangerous, and the wisdom of the crowd fragile.

REFERENCES

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.

Bagdikian, Ben H. 2004. *The New Media Monopoly: A Completely Revised and Updated Edition with Seven New Chapters.* Boston: Beacon Press.

Battaglini, Marco, Rebecca B. Morton, and Thomas R. Palfrey. 2008. "Information Aggregation and Strategic Abstention in Large Laboratory Elections." *American Economic Review* 98 (2): 194–200.

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): 1591–625.

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.

Berger, Roger L. 1982. "Multiparameter Hypothesis Testing and Acceptance Sampling." *Technometrics* 24 (4): 295–300.

Bernardo, Antonio E., and Ivo Welch. 2001. "On the Evolution of Overconfidence and Entrepreneurs." *Journal of Economics and Management Strategy* 10 (3): 301–30.

Bhattacharya, Sourav, John Duffy, and Sun-Tak Kim. 2014. "Compulsory Versus Voluntary Voting: An Experimental Study." *Games and Economic Behavior* 84: 111–31.

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): 107–49.

Camerer, Colin, and Dan Lovallo. 1999. "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review* 89 (1): 306–18.

Danz, David, Lise Vesterlund, and Alistair J. Wilson. Forthcoming. "Belief Elicitation and Behavioral Incentive Compatibility." *American Economic Review*.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315–72.

Elbittar, Alexander, Andrei Gomberg, César Martinelli, and Thomas R. Palfrey. 2020. "Ignorance and Bias in Collective Decisions." *Journal of Economic Behavior and Organization* 174: 339–59.

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): 1623–72.

Feddersen, Timothy J., and Wolfgang Pesendorfer. 1996. "The Swing Voter's Curse." *American Economic Review* 86 (3): 408–24.

Feddersen, Timothy J., and Wolfgang Pesendorfer. 1999. "Abstention in Elections with Asymmetric Information and Diverse Preferences." *American Political Science Review* 93 (2): 381–98.

Fehr, Ernst, and Jean-Robert Tyran. 2005. "Individual Irrationality and Aggregate Outcomes." *Journal of Economic Perspectives* 19 (4): 43–66.

Fischbacher, Urs, 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10 (2): 171–78.

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: 127–50.

Goeree, Jacob K., Thomas R. Palfrey, Brian W. Rogers, and Richard D. McKelvey. 2007. "Self-Correcting Information Cascades." *Review of Economic Studies* 74 (3): 733–62.

Herrera, Helios, Aniol Llorente-Saguer, and Joseph McMurray. 2019a. "The Marginal Voter's Curse." *Economic Journal* 129 (624): 3137–53.

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.

Kahneman, Daniel. 2011. "Don't Blink! The Hazards of Confidence." *New York Times,* October 19.

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.

Kartal, Melis, and Jean-Robert Tyran. 2022. "Replication Data for: Fake News, Voter Overconfidence, and the Quality of Democratic Choice." *American Economic Association* [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E171882V1.

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): 2083–123.

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): 1121–34.

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): 480–98.

Lantian, Anthony, Virginie Bagneux, Sylvain Delouvée, and Nicolas Gauvrit. 2021. "Maybe a Free Thinker but not a Critical One: High Conspiracy Belief Is Associated with Low Critical Thinking Ability." *Applied Cognitive Psychology* 35 (3): 674–84.

Lazer, David, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, et al. 2018. "The Science of Fake News." *Science* 359 (6380): 1094–96.

Levy, Gilat, and Ronny Razin. 2015. "Correlation Neglect, Voting Behavior, and Information Aggregation." *American Economic Review* 105 (4): 1634–45.

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

Marlow, Thomas, Sean Miller, and J. Timmons Roberts. 2021. "Bots and Online Climate Discourses: Twitter Discourse on President Trump's Announcement of US Withdrawal from the Paris Agreement." *Climate Policy* 21 (6): 765–77.

McLennan, Andrew. 1998. "Consequences of the Condorcet Jury Theorem for Beneficial Information Aggregation by Rational Agents." *American Political Science Review* 92 (2): 413–18.

McMurray, Joseph C. 2013. "Aggregating Information by Voting: The Wisdom of the Experts versus the Wisdom of the Masses." *Review of Economic Studies* 80 (1): 277–312.

Mengel, Friederike, and Javier Rivas. 2017. "Common Value Elections with Private Information and Informative Priors: Theory and Experiments." *Games and Economic Behavior* 104 (2017): 190–221.

Moore, Don A., and Daylian M. Cain. 2007. "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): 502–17.

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: 461–81.

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.

Ortoleva, Pietro, and Erik Snowberg. 2015. "Overconfidence in Political Behavior." *American Economic Review* 105 (2): 504–35.

Nyhan, Brendan. 2020. "Facts and Myths about Misperceptions." *Journal of Economic Perspectives* 34 (3): 220–36.

Rubin, Mark. 2021. "When to Adjust Alpha During Multiple Testing: A Consideration of Disjunction, Conjunction, and Individual Testing." *Synthese* 199 (3): 10969–11000.

Serra-Garcia, Marta, and Uri Gneezy. 2021. "Mistakes, Overconfidence, and the Effect of Sharing on Detecting Lies." *American Economic Review* 111 (10): 3160–83.

Ståhl, Tomas, and Jan-Willem Van Prooijen. 2018. "Epistemic Rationality: Skepticism toward Unfounded Beliefs Requires Sufficient Cognitive Ability and Motivation to be Rational." *Personality and Individual Differences* 122: 155–63.

Swami, Viren, Martin Voracek, Stefan Stieger, Ulrich S. Tran, and Adrian Furnham. 2014. "Analytic Thinking Reduces Belief in Conspiracy Theories." *Cognition* 133 (3): 572–85.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–51.