# Strategic Misinformation: The Role of Heterogeneous Confirmation Bias*

Satoshi Kasamatsu† Daiki Kishishita‡

November 24, 2022

**Abstract**

Why do politicians sometimes support evidently false claims, such as fake news and conspiracy theories, spreading among a minority of the electorate? We analyze this political supply of misinformation from the perspective of politicians' electoral incentives. For this purpose, we construct a two-period electoral accountability model in which the incumbent politician decides whether to support or deny the truth about a particular issue and exerts effort for producing public goods. We show that the low-competent incumbent paradoxically has an electoral incentive to support misinformation, even if only a minority of voters believe the misinformation. The key to this paradoxical incentive is the heterogeneity of confirmation bias, a psychological bias that has been regarded as a determinant for people's persistent acceptance of misinformation in the literature. We show that the low-competent politician denies the truth if and only if the variance in the distribution of confirmation bias among voters is sufficiently large. This result indicates that individual differences in psychological bias among voters play a key role in the political supply of misinformation.

**Keywords:** Misinformation; Confirmation bias; Accountability; Reelection; Individual difference

**JEL classification:** D72, D83, D91

*We would like to thank Yu Awaya, Kazumi Hori, Atsushi Kajii, Daisuke Nakajima, Hitoshi Sadakane, Yasuhiro Shirata, Takuro Yamashita, and the participants of the 2022 Summer Workshop on Economic Theory in Otaru for their helpful comments. All remaining errors are our own. Kasamatsu was financially supported by the JSPS KAKENHI Grant Number 22K13422. Kishishita was financially supported by the JSPS KAKENHI Grant Number 22K13368.

†Faculty of Economics, Musashi University. 1-26-1, Toyotamakami, Nerima-ku, Tokyo, Japan 176-0011. E-mail: kasamatsu00@gmail.com

‡School of Management, Tokyo University of Science. 1-11-2, Fujimi, Chiyoda-ku, Tokyo, Japan. 102-0071. E-mail: daiki.kishishita@gmail.com

# 1 Introduction

The current era is characterized by the spread of misinformation, fake news, and conspiracy theories that are evidently false (Jerit and Zhao, 2020; Nyhan, 2020).[1] A notable feature of this spread is that even politicians holding office support false claims. Examples include the former president Donald Trump in the US and the president Jair Bolsonaro in Brazil to name a few.[2] To deal with this prevalence of misinformation among elected politicians,[3] institutions such as PolitiFact in the US conduct fact-checking of their claims. These observations lead us to ask the following question: Why do elected politicians sometimes support arguments that are evidently false? The present study aims to answer this question from the perspective of their electoral incentives.

Because politicians' electoral incentives depend on the demand side, a natural starting point is to understand why voters persistently believe misinformation. The literature on political psychology has extensively analyzed this issue and has shown that psychological biases are key determinants of persistently accepting misinformation. That is, to reach a desired conclusion, people process political information in a biased way in contrast to Bayesian learning, which is called *(directional) motivated reasoning* (see Flynn, Nyhan and Reifler (2017) for a literature review). As a result, even after facing factual information, they may not correct their false hypothesis.[4] For example, Nyhan and Reifler (2010) experimentally find that correcting misinformation that Saddam Hussein had WMD not mitigated but reinforced this misinformation among conservatives. In particular, it has been documented that people have *confirmation bias*, a tendency of "seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand." (Nickerson, 1998, p.175). "To the extent that people are already misinformed, the confirmation bias can clearly facilitate the impact of misinformation on the people who already hold misinformation-consonant beliefs" (Pantazi, Hale and Klein, 2021, p.277).

While these studies show the reason why a non-negligible fraction of voters persistently believe misinformation, this demand-side factor cannot readily explain its political supply side. Even in the presence of psychological biases, it is unlikely that a majority of voters believe a particular type of misinformation. For example, while COVID-19-related conspiracy theories are common across the world, only 31% of the US citizens believe the conspiracy theory that the virus was purposefully created and spread as of 2020 (Uscinski et al., 2020). As such, the

[1]See Pantazi, Hale and Klein (2021) for the discussion about the conceptions of these terms. Unlike misinformation and fake news, not all conspiracy theories are demonstrably false. However, in most cases, they are false. In the following, we focus on the aspect that information is false and use misinformation as an umbrella term.

[2]For information about President Trump, see "A month-by-month look at Donald Trump's top lies of 2019" (CNN News, December 31, 2019, https://edition.cnn.com/2019/12/31/politics/fact-check-donald-trump-top-lies-of-2019-daniel-dale/index.html ). In addition, President Bolsonaro repeatedly argued misinformation about COVID-19 (Ricard and Medeiros, 2020).

[3]Lasser et al. (forthcoming) examine whether members of parliament in the US, UK, and Germany share untrustworthy information on Twitter. They find that a certain share of elected politicians share untrustworthy information, while it is also the case that many politicians share only trustworthy information.

[4]Though the literature has emphasized such biased updating, whether the observed empirical patterns cannot be explained by Bayesian learning is not so straightforward. See Little (2022) for further discussion.

majority of voters are likely to know the truth. If so, as the median voter theorem predicts, politicians should support the truth to win an election. Without putting electoral incentives aside,[5] we aim to resolve this puzzle by emphasizing the overlooked role of heterogeneous confirmation bias among voters.

For this purpose, we extend a canonical two-period political agency model à la Besley (2006) (see Duggan and Martinelli (2017) for a literature review). Our model comprises three types of players: the incumbent politician, a challenger, and a continuum of voters. In period 1, the incumbent politician exerts effort, which influences her performance level regarding public goods provision. After observing the performance level, voters decide whether to reelect the incumbent or elect the challenger. In period 2, the elected politician exerts effort for producing public goods. Politicians are either the ethical type sincerely exerting the highest effort or the opportunistic type who cares for effort cost as well as reelection incentive. Each politician's type is unknown to voters.

The novelty of our model is to introduce a communication stage into this standard model of political agency. At the beginning of the game, players face a controversial issue, for which we take the safety of a COVID-19 vaccine as a leading example. Some voters have a prior that the vaccine may not be safe, while others have a prior that it is likely to be safe. The factual news that the vaccine is safe is reported, but some voters who initially had an anti-vaccine concern misinterpret this news and reinforce their initial argument due to confirmation bias. As a result, some voters remain *misinformed* and deny the truth. Note that we still assume that a majority of voters are *informed* and correctly understand the truth. In this environment, the incumbent politician chooses whether to support the truth reported by the news or deny the truth, which is a cheap-talk message. The sincere type is assumed to tell the truth. Building on Rabin and Schrag (1999), we model voters' confirmation bias such that higher bias increases the probability of misinterpreting news and reinforcing the initial argument when facing belief-inconsonant news.

We show that the opportunistic-type incumbent politician may have an incentive to deny even the truth supported by a majority of voters. In particular, such an incentive may arise when her level of competence in producing public goods is low. This happens neither because she misunderstands the truth nor because she wants to promote a particular ideology. In our model, politicians are assumed to know the truth, and their payoff is simply defined by the benefit of reelection minus the effort cost. Nonetheless, to win an election, the low-competent incumbent may paradoxically deny the truth supported by a majority of voters.

The key to resolving this paradox is the heterogeneity of the degree of confirmation bias among voters. We show that the low-competent incumbent denies the truth if and only if the degree of confirmation bias is sufficiently heterogeneous between voters. Several recent studies of political psychology have analyzed the existence and the causes of individual differences in motivated reasoning such as confirmation bias. For example, Arceneaux and Vander Wielen

[5]A competing hypothesis for resolving the paradox is that politicians deny the truth because they want to do so ideologically without caring about an election. However, Nyhan and Reifler (2015) empirically show that fact-checking reduces politicians' misinformation because they care about the reputation risk, which rejects this hypothesis.

(2017) find that a difference in need for cognition and need for affect determines the level of motivated reasoning, whereas Taber and Lodge (2006) and Kahan et al. (2017) find that a difference in political knowledge and science curiosity correlates with it respectively.[6] Our result shows that the extent of such individual differences in psychological biases creates room for the political supply of denying the truth.

The mechanism is understood as follows. As a benchmark, first, suppose that the degree of confirmation bias is common among voters. Whether to support the truth influences who is the incumbent's political support base. On the one hand, when the incumbent supports the truth, informed voters realize that the incumbent supports the truth, whereas misinformed voters think that the incumbent argues the false statement. As a result, informed voters upwardly update the probability of the incumbent being ethical, but this probability becomes zero for misinformed voters. Hence, her political support base is informed voters when supporting the truth. On the other hand, when the incumbent denies the truth, her reputation becomes zero among informed voters, whereas her reputation is improved among misinformed voters. As such, her political support base is misinformed voters when denying the truth. As we all know, the larger political support base increases the chance of reelection. Since a majority of voters are assumed to be informed, the incumbent never strategically denies the truth, without individual differences in the degree of confirmation bias.

However, a part of this logic fails when the degree of confirmation bias is heterogeneous across voters. The key is that the distribution of confirmation bias becomes divergent between informed and misinformed voters in the presence of individual heterogeneity. Higher confirmation bias increases the probability of remaining misinformed after factual information is provided. Hence, in the presence of heterogeneity, the average degree of confirmation bias among informed voters is expected to be lower than among misinformed voters. As a result, targeting misinformed voters as a political support base enables the incumbent to weaken the electoral accountability mechanism. To see this, note that with higher confirmation bias, voters are more likely to misinterpret the incumbent's low performance in producing public goods as high performance once the incumbent's reputation was established among them. Hence, the higher bias the less likely voters are to punish the incumbent with low performance. This implies that targeting voters with higher bias as the political support base is better for the incumbent.

Taken together, these properties indicate that the incumbent faces both pros and cons of denying the truth in the presence of heterogeneous confirmation bias. Its cost is that her political support base shrinks. Its benefit is that her political support base has a higher confirmation bias so that electoral punishment becomes weaker. For low-competent politicians, their performance in producing public goods tends to be low so that the benefit of denying the truth is large. As a result, the benefit overturns the cost when the degree of confirmation bias is sufficiently heterogeneous between voters. As such, low-competent politicians take a seemingly paradoxical behavior in the presence of individual heterogeneity of confirmation bias: deny the truth supported by a

[6]Relatedly, Enke and Zimmermann (2019) show the individual heterogeneity of correlation neglect, another well-known type of biased updating, and find that the degree of correlation neglect is correlated with cognitive skills.

majority of voters to capture voters with high confirmation bias as the political support base.

Our study not only resolves the puzzle about the political supply of misinformation but also provides several implications for the empirical analysis of misinformation, which will be discussed in Section 4.

**Related literature:** First, our study contributes to the growing literature that theoretically analyzes the effect of voter confirmation bias on elections (Levy and Razin, 2015; Lockwood, 2017; Millner, Ollivier and Simon, 2020; Little, Schnakenberg and Turner, 2022).[7] Particularly, Lockwood (2017) finds its negative effect on electoral accountability based on the setting of Rabin and Schrag (1999) like ours;[8] and Little, Schnakenberg and Turner (2022) develop a general framework describing motivated reasoning and apply it to the analysis of electoral accountability. While our study is motivated by these studies, neither of them analyzes misinformation nor heterogeneity of confirmation bias.[9]

Second, the literature on propaganda examines whether politicians can induce a particular action of voters by manipulating information (e.g., Glaeser, 2005; Gehlbach and Sonin, 2014; Little, 2017; Guriev and Treisman, 2020; Horz, 2021). These studies typically consider a situation where an authoritarian regime manipulates information about the quality of the regime to prevent protests. Hence, what politicians manipulate is the information directly related to their office-holding or policy goals, which is also applicable to certain situations in democracy (e.g., when a politician losing an election supports false claims about electoral fraud). On the contrary, our model assumes that whether to deny the truth is not directly relevant to reelection or policymaking. Nonetheless, politicians deny the truth because it indirectly influences reelection through the effect on electoral accountability. We complement the findings of this literature by presenting another incentive for misinformation.[10]

Third, denying the truth is unpopular in our model in that a majority of voters do not support it. Therefore, our study is also related to studies showing that unpopular policies may paradoxically increase the chance of electoral winning. The mechanisms proposed by the existing studies include their roles of cultivating core supporters (Glaeser, Ponzetto and Shapiro, 2005), serving as a commitment to invest in the quality of policymaking (Carrillo and Castanheira, 2008), serving as a signal of the politician being ethical (Acemoglu, Egorov and Sonin, 2013), and investing an issue that may become popular in the future (Eguia and Giovannoni, 2019). We contribute to this literature by proposing a new mechanism induced by

[7]Subsequent works on modeling confirmation bias include Yariv (2002), Fryer, Harms and Jackson (2019), and Wilson (2014). Motivated reasoning includes other types of biased updating such as wishful thinking (Bénabou and Tirole, 2016). Le Yaouanq (2021) analyzes the effect of wishful thinking on direct democracy.

[8]He analyzes two models of political agency: one is where the optimal policy and the performance are observable such as the incumbent's effort choice, and the other one is where these are unobservable. For the latter model, he shows that confirmation bias reduces the incumbent's pandering incentive.

[9]Kartal and Tyran (2022) analyze the effect of overconfidence and misinformation on information aggregation in a common-interest voting game. In addition to a difference in psychological bias they focus on, politicians' incentives are not examined in their model because it concerns the efficiency of direct democracy.

[10]Another distinctive feature is that misinformation is already spreading among a part of voters in our model. Hence, what a politician does is not to invent misinformation, but to support existing misinformation.

confirmation bias. The work of Grillo and Prato (forthcoming) is particularly related. They show that even when most citizens value democracy, democratic backsliding can occur in case where voters have reference-dependent preferences. The mechanism is that, with reference dependence, the incumbent's good performance after backsliding is highly evaluated by voters because of its surprisingness. Like ours, an unpopular policy could be beneficial for reelection in the presence of psychological biases because it changes voters' response to the incumbent's performance. The novelty of our study is to show the role of heterogeneous psychological biases.

Lastly, the most related study is that by Szeidl and Szucs (2022), which analyzes why a politician persuades voters of a conspiracy theory. Their model assumes that voters believing the conspiracy theory come to distrust the intellectual elite and not to believe the media's news. Hence, in order to weaken electoral punishment based on news reporting, the incumbent with low quality has an incentive to persuade voters of the conspiracy theory. Our model has three distinctive features. First, the aspect of misinformation we consider is that it is just false, whereas their focus is the aspect that it is often tied to the perception that elites such as the mainstream media are corrupt. Second, we consider confirmation bias, an important bias that has attracted attention in the literature on political psychology, whereas their model does not contain it. Third, politicians deny the truth even if a majority correctly understands the truth in our model, while their model does not predict it because their model contains only a decisive voter.[11]

# 2 Model

Our model is an extension of a standard two-period electoral accountability à la Besley (2006). It comprises three types of players: the incumbent politician (we use female pronoun), a challenger, and a continuum of voters (we use male pronoun). In period 1, the incumbent politician chooses her effort level for public goods provision. Based on the incumbent's performance, voters vote for either the incumbent or the challenger. The candidate obtaining a majority of votes holds the office and chooses her effort level in period 2. By introducing a communication stage, where the incumbent communicates her opinion about a controversial issue, at the beginning of this model, we analyze the conditions under which the incumbent expresses misinformation about the controversial issue. The timing of the game is visualized in Figure 1.

## 2.1 Communication Stage

There is a controversial issue whose truth is uncertain, which, for example, can be interpreted as whether taking a COVID-19 vaccine is safe. More formally, the truth is given by $\omega \in \{0, 1\}$. While $\omega$ is uncertain for voters, the incumbent knows its value.

[11]Relatedly, Grossman and Helpman (2022) analyze the strategic use of misinformation about the policy environment and parties' positions in a model of electoral competition. They show that policy divergence can result when parties can persuade a part of their partisans to believe false arguments. In their model, parties' use of misinformation does not undermine the popularity among informed voters so that using misinformation is always beneficial for parties. In contrast, it can improve the reelection probability in our model even if it reduces the popularity among a majority of voters.

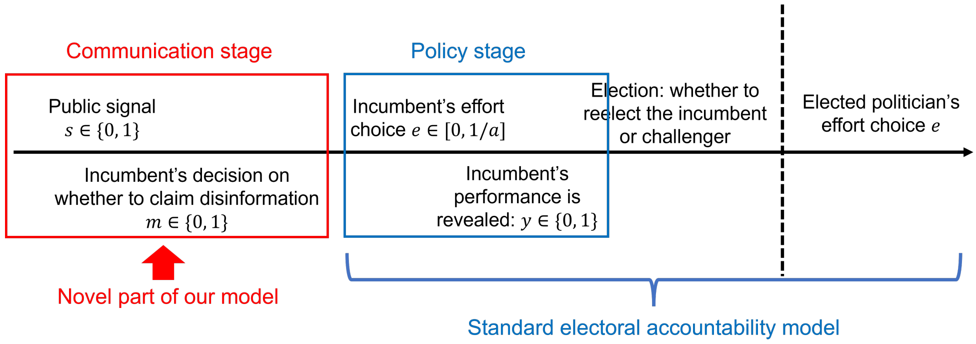


Figure 1: Timing of the Game

**Heterogeneous prior:** Voters have heterogeneous priors on the probability of $\omega$ being one. Let the true value of $\omega$ be $k$. We suppose that a fraction $\kappa$ of voters has a prior of $\Pr(\omega = k) = \bar{\alpha} \in (0.5, 1)$, whereas a fraction $1 - \kappa$ of voters has a prior that $\Pr(\omega = k) = \underline{\alpha} \in (0, 0.5)$. Each voter does not know the distribution of priors because otherwise, they can infer the value of $\omega$ from the distribution.

**Public signal:** At the beginning of the game, voters receive a public signal about the truth: $s \in \{0, 1\}$. In the example of the safety of the COVID-19 vaccination, $s$ can be interpreted as an official statement by medical experts such as the World Health Organization (WHO) or the media's news reporting the result of clinical trials. To analyze misinformation, we have to consider a situation where the truth is evident under rational reasoning. Otherwise, the truth is objectively uncertain and any belief can be factual even after rational reasoning. Hence, we assume that the public signal $s$ is perfect; that is, $s = \omega$ with probability one. This implies that Bayesian rational agents should learn the truth after the public signal.

**Political communication:** After the public signal, the incumbent sends a cheap-talk message about $\omega$: $m \in \{0, 1\}$. In particular, $m = \omega$ (resp. $m \neq \omega$) represents confirming the truth (resp. denying the truth). Our interest is under what conditions the incumbent sends $m \neq \omega$ to voters. We assume that politicians support the truth in an indifferent case. This is just for simplicity and does not affect our results.

## 2.2 Stage of Public Goods Provision

In period $t(= 1, 2)$, the politician holding the office exerts effort for producing public goods: $e_t \in [0, 1/a]$, where $a > 0$. In period 1, this is after the communication stage. The effort level influences whether public goods are successfully provided (i.e., the elected politician's performance). In particular, the elected politician's performance is given by $y_t \in \{0, 1\}$, where $y_t = 1$ represents that public goods are successfully provided (i.e, high performance). The probability of $y = 1$ is given by $z(e) := ae$. Hence, the higher effort increases the probability of public goods provision, and the performance is always high when the exerted effort is maximum (i.e., $e_t = 1/a$).

## 2.3 Politicians

Politicians are divided into two types: *ethical type* and *opportunistic type*.

The ethical type is a sincere politician in that she truthfully confirms the truth (i.e., $m = \omega$) and exerts effort for public goods provision as much as possible (i.e., $e = 1/a$).

The opportunistic type is a self-interested politician. With probability $1 - \varepsilon \in (0, 1)$, she is strategic and maximizes the following payoff:

$$\sum_{t=1}^{2} \mathbf{1}_t \left( b - \frac{e_t^2}{2\delta} \right),$$

where $\mathbf{1}_t$ is an indicator function that takes one if and only if the politician is holding the office in period $t$. Here, $b > 0$ represents the office-holding benefit, while $e_t^2/\delta$ represents the effort cost for public goods provision. $\delta$ is the politician's productivity or competence because higher $\delta$ reduces the effort cost. We allow $\delta$ to vary between politicians. In particular, $\delta$ is a stochastic variable that follows a uniform distribution $U(0, \Delta)$, where $\Delta > 0$. Our main interest is this politician. To exclude implausible off-path beliefs, we also assume that with an arbitrarily small but positive probability $\varepsilon$, she non-strategically denies the truth and exerts no effort for public goods provision (i.e., $m \neq \omega$ and $e = 0$).

## 2.4 Election

Voters do not know each politician's type, whereas each politician knows her own type. Note that the value of $\delta$ is also known to the politician but unknown to voters. For simplicity, we assume that the prior probability of a politician being ethical is 0.5.

**Voting behavior:** After observing the incumbent's performance in period 1, voters face an election at the beginning of period 2. Voters' payoff from public goods provision in period $t$ is $y_t$. Hence, voters' expected payoff from public goods provision in period 2 when reelecting the incumbent (resp. electing a challenger) is $\mathbb{E}[y_2 \mid y_1, \text{Incumbent}]$ (resp. $\mathbb{E}[y_2 \mid y_1, \text{Challenger}]$).

We suppose probabilistic voting. That is, voter $i \in [0, 1]$ votes for the incumbent politician if and only if

$$\mathbb{E}[y_2 \mid y_1, \text{Incumbent}] + \zeta_i + \eta \geq \mathbb{E}[y_2 \mid y_1, \text{Challenger}]. \tag{1}$$

In addition to the expected utility derived from public goods provision, various stochastic factors influence voting behavior. For example, each voter's decision might be affected by the individual and aggregate level of the economic situation, both of which are stochastic. Following the vast literature on probabilistic voting (e.g., Lindbeck and Weibull, 1987; Dixit and Londregan, 1996), we formulate these factors using $\zeta_i$ and $\eta$. $\zeta_i$ is a shock specific to voter $i$, which follows a uniform distribution $U[-1/(2\gamma), 1/(2\gamma)]$, whereas $\eta$ is an aggregate shock that is common across voters, which follows a uniform distribution $U[-1/(2\psi), 1/(2\psi)]$. Note that $\gamma \in \left(0, \frac{\psi}{2\psi+1}\right]$ and $\psi \in (0, 0.5)$ are assumed to guarantee that their votes are always probabilistic.

Given this setting, voter $i$'s voting behavior is derived as follows. Let $p_i$ be voter $i$'s subjective probability of the incumbent being ethical at the time of the election. In period 2, the opportunistic type chooses the lowest level of effort for public goods provision ($e_2 = 0$) because there is no reelection concern, whereas the ethical type chooses the highest effort level ($e_2 = 1/a$). Hence, voter $i$ votes for the incumbent if and only if

$$\begin{aligned}(17) \Leftrightarrow & p_i + \zeta_i + \eta \geq \frac{1}{2} \\ \Leftrightarrow & \zeta_i \geq \frac{1}{2} - \eta - p_i.\end{aligned}$$

**Probability of winning:** By aggregating each voter's voting behavior, we derive the incumbent's winning probability. Let the expected value of $p_i$ among the entire electorate given $(m, y_1)$ be $\mathbb{E}[p_i \mid m, y_1]$. Then, given $\eta$, the fraction of voters who vote for the incumbent is written as

$$F(\eta) := \frac{1}{2} + \gamma\left(\mathbb{E}[p_i \mid m, y_1] - \frac{1}{2} + \eta\right).$$

Hence, the probability of the incumbent obtaining a majority of votes given $(m, y_1)$ is

$$\Pr\left(F(\eta) \geq \frac{1}{2}\right) = \frac{1}{2} + \psi\left(\mathbb{E}[p_i \mid m, y_1] - \frac{1}{2}\right).$$

Furthermore, $y_1$ is a stochastic variable whose probability of taking one is $ae_1$. Thus, the expected probability of the incumbent obtaining a majority of votes given $(m, e_1)$ is

$$\frac{1}{2} + \psi\left(\mathbb{E}[p_i \mid m, e_1] - \frac{1}{2}\right),$$

where $\mathbb{E}[p_i \mid m, e_1] = ae_1\mathbb{E}[p_i \mid m, y_1 = 1] + (1 - ae_1)\mathbb{E}[p_i \mid m, y_1 = 0]$. This implies that the electoral result only depends on the expected average reputation of the incumbent, $\mathbb{E}[p_i \mid m, e_1]$.

## 2.5 Confirmation Bias

The last, yet most important ingredient is voters' confirmation bias, which is modeled based on the formulation by Rabin and Schrag (1999). In our model, in response to the incumbent's message and the incumbent's performance, voters update their beliefs on two uncertain states: one is the value of $\omega$ and the other one is the incumbent's type.

**Confirmation bias on $\omega$:** Suppose that voter $i$ receives a binary signal $n \in \{0, 1\}$ about $\omega$. This corresponds to either the public signal $s$ or the incumbent's message $m$. Let voter $i$'s subjective belief on the probability of $\omega$ being one just before the arrival signal $n$ be $\alpha$. Let also voter $i$'s subjective belief on the probability of the incumbent's type being ethical just before the arrival of signal $n$ be $\beta$.

To model confirmation bias, we suppose that voters may misinterpret a signal contradicting the current belief. Suppose that voter $i$ perceives that the received signal is $\tilde{n} \in \{0, 1\}$. We allow that $\tilde{n} \neq n$ could happen when signal $n$ contradicts his current hypothesis. The detailed definition of confirmation bias is given as follows:

**Definition 1.** *Suppose that the equilibrium strategies imply that* $\Pr(n = 1 \mid \omega = 1) > 0.5$.[12] *Then, voter i having confirmation bias with degree* $q_i^\omega$ *updates the belief about* $\omega$ *as follows:*

*(i). Suppose that* $\alpha > 0.5$. *When* $n = 1$, *voter i correctly perceives that he receives signal 1 (i.e.,* $\tilde{n} = 1$*). When* $n = 0$, *voter i misperceives that he receives signal 1 (i.e.,* $\tilde{n} = 1$*) with probability* $q_i$*; and correctly perceives that he receives signal* 0 *(i.e.,* $\tilde{n} = 0$*) with the remaining probability.*

*(ii). Suppose that* $\alpha < 0.5$. *When* $n = 0$, *voter i correctly perceives that he receives signal 0 (i.e.,* $\tilde{n} = 0$*). When* $n = 1$, *voter i misperceives that he receives signal 0 (i.e.,* $\tilde{n} = 0$*) with probability* $q_i$*; and correctly perceives that he receives signal* 1 *(i.e.,* $\tilde{n} = 1$*) with the remaining probability.*

*(iii). Suppose that* $\alpha = 0.5$. *Voter i correctly perceives that he receives signal n (i.e.,* $\tilde{n} = n$*).*

*(iv). Voter i does not know a possibility of misperception. That is, he updates the belief on the value of* $\omega$ *based on priors* $\alpha$ *and* $\beta$ *in the Bayesian way as if* $n = \tilde{n}$.

Suppose that voter $i$ thinks that state 1 is more likely to be true than state 0. Since signal 1 is consistent with his belief, he does not misperceive this signal. However, when he receives signal 0 that contradicts his prior, he misperceives that he receives the opposite signal (i.e., signal 1) with probability $q_i^\omega$,[13] which is (i) in the definition. In addition, when he thinks that state 0 is more likely to be true, the updating is defined symmetrically (see (ii)). Lastly, when he thinks that both states are equally likely to be true, his prior is neutral so that he never misperceives the signal (see (iii)). When $q_i^\omega = 0$, the situation is reduced to be Bayesian updating; thus, the value of $q_i^\omega$ can be interpreted as the degree of confirmation bias. Note that each voter is naive in

[12]Which signal suggests state 1 depends on the incumbent's equilibrium strategy. For example, which message indicates state 1 depends on what message the incumbent sends in the equilibrium. See Lockwood (2017) for the detailed discussion.

[13]With a positive probability, he interprets a belief-inconsonant signal as a belief-consonant signal rather than ignoring it. As a result, his confidence in his prior hypothesis increases despite the the information suggesting the opposite. While it seems odd at first glance, this is indeed consistent with experimental findings. First, it has been documented that people's beliefs could diverge after observing the same piece of information (e.g., Lord, Ross and Lepper, 1979; Fryer, Harms and Jackson, 2019). Second, several studies have found the backfire effects that correcting misinformation rather increases misperceptions (e.g., Nyhan and Reifler, 2010) (to be fair, results are mixed across studies (Swire-Thompson, DeGutis and Lazer, 2020)). The mechanism is understood as follows. First, corrective information is typically reported as news pitting two sides of an argument against each other. In this situation, agents who disregard arguments contradicting their opinion hold a more extreme opinion because they receive only a belief-consonant argument included in the news (e.g., Nyhan and Reifler, 2010; Fryer, Harms and Jackson, 2019). Second, when receiving information contradicting their opinion, agents may generate counter-arguments by spending time and cognitive cost, which reinforces their initial opinion (e.g., Taber and Lodge, 2006; Taber, Cann and Kucsova, 2009).

the sense that he does not notice that he is suffering from confirmation bias, which is condition (iv).[14]

**Confirmation bias on the incumbent's type:** Let $t^I \in \{e, o\}$ being the incumbent's type where $e$ represents the incumbent is ethical. Then, as in the above, confirmation bias is defined:

**Definition 2.** *Suppose that the equilibrium strategies imply that* $\Pr(n = 1 \mid t^I = e) > 0.5$. *Then, voter $i$ having confirmation bias with degree $q_i^t$ updates the belief about $t^I$ as follows:*

*(i). Suppose that $\beta > 0.5$. When $n = 1$, voter $i$ correctly perceives that he receives signal 1 (i.e., $\tilde{n} = 1$). When $n = 0$, voter $i$ misperceives that he receives signal 1 (i.e., $\tilde{n} = 1$) with probability $q_i$; and correctly perceives that he receives signal* 0 *(i.e., $\tilde{n} = 0$) with the remaining probability.*

*(ii). Suppose that $\beta < 0.5$. When $n = 0$, voter $i$ correctly perceives that he receives signal 0 (i.e., $\tilde{n} = 0$). When $n = 1$, voter $i$ misperceives that he receives signal 0 (i.e., $\tilde{n} = 0$) with probability $q_i$; and correctly perceives that he receives signal* 1 *(i.e., $\tilde{n} = 1$) with the remaining probability.*

*(iii). Suppose that $\beta = 0.5$. Voter $i$ correctly perceives that he receives signal $n$ (i.e., $\tilde{n} = n$).*

*(iv). Voter $i$ does not know a possibility of misperception. That is, he updates the belief on the value of $\omega$ based on prior $\alpha$ and $\beta$ in the Bayesian way as if $n = \tilde{n}$.*

While our modeling is basically the same as that of Rabin and Schrag (1999), a distinctive feature is that voters face multi-dimensional uncertainty in our model. To account for this, our definition assumes that voter $i$ updates the belief about $\omega$ and the incumbent's type *separately*; that is, he separately updates the marginal probability instead of updating the joint probability because updating a full joint distribution is cognitively taxing (Loh and Phelan, 2019). To see how this assumption works, suppose that voter $i$ receives signal $n$ suggesting that $\omega = 1$ and the incumbent is ethical. When $\alpha < 0.5$, the voter may misinterpret signal 1 as signal 0 in updating about $\omega$. We assume that this does not influence updating about the incumbent's type: $\tilde{n}$ could be different between updating about $\omega$ and updating about the incumbent's type.[15]

**Heterogeneity:** We introduce individual differences in the degree of confirmation bias. We suppose that $q_i^\omega$ and $q_i^t$ follow a common distribution $F$ (either continuous or discrete), whose

[14] Whether each voter is aware of other voters' confirmation bias does not matter because his voting decision solely depends on the incumbent's equilibrium strategy and the equilibrium strategies of the other voters is irrelevant.

[15] Otherwise, defining confirmation bias is not straightforward, which can be seen in the following two situations. First, suppose that $\alpha < 0.5$ and $\beta < 0.5$. Then, in both dimensions, voter $i$ has a tendency to misinterpret signal 1, but its probability differs ($q_i^\omega$ vs. $q_i^t$). If we assume common $\tilde{n}$ across different issues, we need to define the probability of misinterpretation as a combination of $q_i^\omega$ and $q_i^t$. Second, suppose that $\alpha < 0.5$ and $\beta > 0.5$. Then, signal 1 contradicts the voter's hypothesis on $\omega$, but it is consistent with the voter's hypothesis on the incumbent's type. If we assume common $\tilde{n}$ across different issues, we need to introduce an ad-hoc assumption about which hypothesis is prioritized. Our assumption that voters update the marginal belief separately avoids these issues.

variance is given by $\sigma^2$. Note that $q \in [0, 1]$, and thus the upper bound of the variance exists. A straightforward calculation yields that it is $\mathbb{E}[q](1 - \mathbb{E}[q])$. Higher $\sigma^2$ represents larger heterogeneity in the degree of confirmation bias across voters.

While $\sigma^2$ is our main interest, another dimension of heterogeneity could be within individuals. That is, $q_i^\omega$ and $q_i^t$ may vary for the same voter because the issue is different. To describe this, we suppose that $q_i^\omega = q_i^t$ are the same for all $i$ with probability $\lambda$, but $q_i^\omega$ and $q_i^t$ are independently drawn from $F$ for all $i$ with the remaining probability. In this setting, $\lambda \in (0, 1]$ captures how correlated confirmation bias is across issues. We also examine this type of heterogeneity to see that heterogeneity across individuals has a distinctive feature.

### 2.6 Timing of the Game and Equilibrium Concept

The timing of the game is given as follows:

1. Voters observe a public signal about $\omega$: $s \in \{0, 1\}$.

2. The incumbent sends a message about $\omega$: $m \in \{0, 1\}$. In addition, she exerts effort for producing public goods: $e_1 \in [0, 1/a]$, which is unobservable to voters.

3. Voters observe whether public goods are successfully provided (i.e., the incumbent's performance level): $y_1 \in \{0, 1\}$. Then, they vote for the incumbent or the challenger.

4. The elected politician exerts effort, $e_2 \in [0, 1/a]$, and the performance level is realized.

The equilibrium concept is pure-strategy perfect Bayesian equilibrium where the belief updating is biased due to confirmation bias. As in various incomplete information games, our equilibrium concept allows off-path beliefs to be unreasonable. To restrict our attention to reasonable ones, we assume that each voter believes that the incumbent is opportunistic if he thinks that an off-path is observed.

## 3 Equilibrium

Because the situation is symmetric independently of the value of $\omega$, hereafter, without loss of generality, we focus on the case where $\omega = 1$. Hence, $m = 1$ represents confirming the truth, whereas $m = 0$ represents denying the truth.

### 3.1 Belief Polarization after the Public Signal

At the beginning of the game, voters receive a public signal $s = 1$ because $\omega$ is assumed to be one. While they observe the common signal, a certain fraction of voters misperceive the signal due to their confirmation bias. Consequently, public opinion on the value of $\omega$ becomes polarized.

Let $\tilde{s}_i$ be the perceived signal voter $i$ receives and $w_i$ be voter $i$'s subjective probability of $\omega$ being one given $\tilde{s}_i$. First, consider the $\kappa$ fraction of voters who initially believed that state 1 is more likely to be true. For them, signal 1 confirms their initial hypothesis so that they perceive signal 1 literally. Hence, for all of them, $w_i = 1$.

Next, consider $1-\kappa$ fraction of voters who initially believed that state 0 is more likely to be true. For them, signal 1 contradicts their initial hypothesis so that they may misperceive signal 1 as signal 0. In particular, voter $i$ misperceives the signal and believes that the truth is $\omega = 0$ (i.e., $w_i = 0$) with probability $q_i^{\omega}$, whereas he literally perceives signal 1 and learns the truth (i.e., $w_i = 1$) with probability $1 - q_i^{\omega}$. Hence, among the $1-\kappa$ fraction of voters, $\mathbb{E}[q]$ fraction of them believes the misinformation, whereas $1 - \mathbb{E}[q]$ fraction of them believes the truth.

Consequently, despite that voters observe the same public signal, their beliefs diverge more due to confirmation bias. This belief polarization after observing the same piece of information looks at odds at first glance, but it is indeed consistent with empirical findings. For example, the seminal experiment by Lord, Ross and Lepper (1979) assigned to subjects research summaries about the deterrent efficacy of the death penalty. They found that subjects' beliefs polarized despite that they read the same piece of information.[16]

In summary, we obtain the following lemma:

**Lemma 1.** *After the public signal, $\kappa + (1-\kappa)(1-\mathbb{E}[q])$ fraction of voters are informed in that they firmly believe the truth (i.e., $w_i = 1$). On the other hand, the remaining fraction of voters are misinformed in that they firmly believe the misinformation (i.e., $w_i = 0$).*

If a majority of voters are misinformed, it is nothing strange that the incumbent supports the misinformation. Our interest is whether the incumbent confirms the misinformation even if only a minority of voters believe the false. To analyze this interesting case, throughout the paper, we assume the following:

**Assumption 1.** $\kappa + (1-\kappa)(1-\mathbb{E}[q]) > \frac{1}{2}$ *holds.*

## 3.2 Incumbent's Message and Political Support Base

After the public signal induces belief polarization, the incumbent decides whether to support or deny the truth. This message influences who are the incumbent's political support base. To see the details, let the equilibrium message of the opportunistic type with $\delta$ when she is strategic and the state is $\omega$ be $m^*(\delta,\omega)$ and let the expected value of $m^*(\delta,\omega)$ over $\delta$ be $\mathbb{E}[m^*(\delta,\omega)] := \int_0^{\Delta} \frac{m^*(\delta,\omega)}{\Delta} d\delta$. In addition, with probability $\varepsilon$, the opportunistic type is non-strategic and always denies the truth. Hence, the opportunistic type confirms the truth with probability $(1-\varepsilon)\mathbb{E}[m^*(\delta,\omega)]$. Without loss of generality, we assume symmetric strategies such that whether to confirm the truth is independent of $\omega$ i.e., $m^*(\delta, 1) = 1 - m^*(\delta, 0)$.

[16]Various subsequent works support their finding (e.g., Taber and Lodge, 2006; Taber, Cann and Kucsova, 2009; Fryer, Harms and Jackson, 2019). Note that this does not imply that such belief polarization can never be explained by Bayesian learning. In a rich setting such as with multidimensional uncertainty, even Bayesian learners may face belief polarization (e.g., Andreoni and Mylovanov, 2012).

As seen above, informed and misinformed voters believe a different state as the truth. Hence, the evaluation of the incumbent's message differs between informed and misinformed voters. To see this, let voter $i$'s subjective probability of the incumbent being ethical given $m$ be $p_I^{int}(m)$ (resp. $p_M^{int}(m)$) when voter $i$ is informed (resp. misinformed). This is the interim belief in that the belief will be further updated when the incumbent's performance is realized. Hence, to distinguish it from the belief at the time of the election, $p$, we use $p^{int}$.

**Lemma 2.**

*(i). Informed voters evaluate the incumbent as follows:*

$$p_I^{int}(1) = \bar{p} := \frac{1}{1 + (1 - \varepsilon)\mathbb{E}[m^*(\delta, 1)]} > 0.5; \;\; p_I^{int}(0) = 0.$$

*(ii). Misinformed voters evaluate the incumbent as follows:*

$$p_M^{int}(1) = 0; \;\; p_M^{int}(0) = \bar{p} > 0.5.$$

The initial probability of the incumbent being ethical is half so that voters have no hypothesis about whether the incumbent is likely to be ethical or not. Hence, voters update their belief on the incumbent's type in the Bayesian manner; that is, voters literally perceive the incumbent's message.[17] The updated beliefs are summarized in the above lemma.

First, consider informed voters who correctly believe that $\omega = 1$. For them, message 1 is regarded as confirming the truth so that they upwardly update the incumbent's reputation after observing message 1. On the contrary, message 0 means denying the truth, which is never chosen by the ethical type. Hence, the incumbent's reputation becomes zero among informed voters.

Second, consider misinformed voters who falsely believe that $\omega = 0$. For them, message 1 is regarded as denying the truth, though its actual meaning is to support the truth. Hence, the incumbent's reputation becomes zero among misinformed voters. On the contrary, for them, message 0 is regarded as confirming the truth, though its actual meaning is to deny the truth. Without noticing the actual meaning, they upwardly update the incumbent's reputation.

Once the reputation becomes zero, it will not be restored again. Hence, this result implies that whether to confirm the truth determines who is the incumbent's political support base. If the incumbent confirms the truth, her political support base is informed voters, while her political support base is misinformed voters if the incumbent denies the truth.

From Assumption 1, a majority of voters are assumed to be informed. Hence, from the perspective of maximizing the size of the political support base, the incumbent should not deny the truth. If she does so, her reputation becomes zero among a majority of voters. Later, we

[17]In practice, a certain fraction of voters initially have either strong anti- or pro-incumbent beliefs. We ignore them in our analysis because their voting choice would be deterministic independent of the incumbent's message. Indeed, Swire-Thompson et al. (2020) experimentally show that correcting misinformation by a politician does not change the evaluation toward the politician among subjects having pro- or anti-sentiment toward the politician. Since an electoral result is likely to be determined by swing voters, it suffices to focus on such decisive voters initially having a neutral belief on the incumbent.

prove that the opportunistic type may deny the truth in spite of this cost of doing so even if she is strategic.

## 3.3 Reelection Probability

The reelection probability when the incumbent chooses message $m$ and exerts effort $e_1$ is given by

$$P^*(m, e_1) = \frac{1}{2} + \psi\left(\mathbb{E}[p_i \mid m, e_1] - \frac{1}{2}\right),$$

implying that it is increasing in the incumbent's expected average reputation at the time of the election.

**Notations:** To derive the value of $\mathbb{E}[p_i \mid m, e_1]$, we need to define the equilibrium probability of the opportunistic type's performance being high. Otherwise, we cannot compute the updated belief given the perceived level of the incumbent's performance.

On the one hand, with probability $1-\varepsilon$, the opportunistic type is strategic and chooses the effort level maximizing her payoff. Let this equilibrium effort in period 1 given $(\delta, m)$ be $e^*(\delta, m)$. On the other hand, with the remaining probability, the opportunistic type non-strategically exerts no effort. Because the probability of high performance is $ae$, these considerations imply that the equilibrium probability of the opportunistic type's performance being high is given by

$$\mathbb{E}[y_o^* \mid m] := \begin{cases} a\int_0^\Delta e^*(\delta, \omega)\frac{m^*(\delta,\omega)}{\Delta\mathbb{E}[m^*(\delta,\omega)]}d\delta & (m = \omega) \\ (1-\varepsilon)a\int_0^\Delta e^*(\delta, m \neq \omega)\frac{1-m^*(\delta,\omega)}{\Delta(1-\mathbb{E}[m^*(\delta,\omega)])}d\delta & (m \neq \omega) \end{cases}.$$

Based on this notation, we derive the value of $\mathbb{E}[p_i \mid m, e_1]$ step by step.

**Incumbent's reputation for voter $i$:** Our starting point is to derive the individual value of $p_i$, which depends on whether voter $i$ perceives that the incumbent's performance is high or not. Let $\tilde{y}_i$ be the incumbent's period 1 performance perceived by voter $i$. This value may differ from the true performance, $y$, due to confirmation bias. Let informed (resp. misinformed) voter $i$'s subjective probability of the incumbent being ethical given $(m, \tilde{y}_i)$ be $p_{Ii}(m, \tilde{y}_i)$ (resp. $p_{Mi}(m, \tilde{y}_i)$). We aim to derive $\mathbb{E}[p_{Ii}(m, \tilde{y}_i) \mid q_i, e]$ and $\mathbb{E}[p_{Mi}(m, \tilde{y}_i) \mid q_i, e]$.

The values of $p_i$ differ depending on whether voter $i$ thinks that the incumbent supports the truth in the communication stage. First, suppose that voter $i$ thinks that the incumbent supports the truth; that is, voter $i$ is in the incumbent's political support base. This corresponds to the situation where $m = 1$ for informed voters and the situation where $m = 0$ for misinformed voters. In this case, the incumbent's reputation is updated as follows given $\tilde{y}_i$:

$$p_{Ii}(1, 1) = p_{Mi}(0, 1) = \frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}; \quad p_{Ii}(1, 0) = p_{Mi}(0, 0) = 0.$$

Note that the ethical type chooses the highest effort level so that her performance is always high.

Now, due to confirmation bias, voter $i$ misperceives low performance as high performance with probability $q_i^t$. Hence, with probability $ae_1 + q_i^t(1-ae_1)$, $\tilde{y}_i = 1$. Therefore, we have

$$\mathbb{E}[p_{Ii}(1,\tilde{y}_i) \mid q_i, e_1] = \mathbb{E}[p_{Mi}(0,\tilde{y}_i) \mid q_i, e_1] = [ae_1 + q_i^t(1-ae_1)]\frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}. \tag{2}$$

This is increasing in $q_i^t$; that is, higher confirmation bias increases the expected reputation of the incumbent.

Next, suppose that voter $i$ thinks that the incumbent denies the truth in the communication stage. This corresponds to the situation where $m = 0$ for informed voters and the situation where $m = 1$ for misinformed voters. Since the reputation is initially zero, it is never restored. That is, regardless of $\tilde{y}_i$,

$$p_{Ii}(0,\tilde{y}_i) = p_{Mi}(1,\tilde{y}_i) = 0.$$

Therefore,

$$\mathbb{E}[p_{Ii}(0,\tilde{y}_i) \mid q_i, e_1] = \mathbb{E}[p_{Mi}(1,\tilde{y}_i) \mid q_i, e_1] = 0. \tag{3}$$

**Incumbent's reelection probability:** Having the derived incumbent's reputation for each voter in hand, we can obtain the incumbent's expected average reputation, $\mathbb{E}[p_i \mid m, e_1]$. By utilizing (2) and (3), we have

$$\begin{aligned}\mathbb{E}[p_i \mid 1, e_1] &= [\kappa + (1-\kappa)(1-\mathbb{E}[q])]\,\mathbb{E}[p_{Ii}(1,\tilde{y}_i) \mid e_1]\\ &= [\kappa + (1-\kappa)(1-\mathbb{E}[q])]\,\frac{\left[ae_1 + \mathbb{E}[q_i^t \mid I](1-ae_1)\right]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]};\end{aligned} \tag{4}$$

$$\begin{aligned}\mathbb{E}[p_i \mid 0, e_1] =&(1-\kappa)\mathbb{E}[q]\mathbb{E}[p_{Mi}(0,\tilde{y}_i) \mid e_1]\\ =&(1-\kappa)\mathbb{E}[q]\frac{\left[ae_1 + \mathbb{E}[q_i^t \mid M](1-ae_1)\right]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]},\end{aligned} \tag{5}$$

where $\mathbb{E}[q_i^t \mid I]$ (resp. $\mathbb{E}[q_i^t \mid M]$) is the average value of $q_i^t$ among informed (resp. misinformed) voters. Therefore, we obtain the reelection probability of the incumbent:

**Lemma 3.** *When the incumbent chooses message $m$ and exerts effort $e_1 = e$, her reelection probability is:*

$$P^*(1,e) := \frac{1-\psi}{2} + \psi\,[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\,\frac{\left[ae + \mathbb{E}[q_i^t \mid I](1-ae)\right]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]};$$

$$P^*(0,e) := \frac{1-\psi}{2} + \psi(1-\kappa)\mathbb{E}[q]\frac{\left[ae + \mathbb{E}[q_i^t \mid M](1-ae)\right]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}.$$

*Hence, the following holds:*

*(i). $P^*(m,e)$ is increasing in $e$ for any $m$.*

*(ii). $P^*(1,e)$ is increasing in $\mathbb{E}[q_i^t \mid I]$, whereas $P^*(0,e)$ is increasing in $\mathbb{E}[q_i^t \mid M]$.*

*(iii). $\frac{\partial P^*(1,e)}{\partial e}$ is decreasing in $\mathbb{E}[q_i^t \mid I]$, whereas $\frac{\partial P^*(0,e)}{\partial e}$ is decreasing in $\mathbb{E}[q_i^t \mid M]$.*

This lemma provides three important observations. (i) shows that $P^*(m,e)$ is increasing in $e$; that is, higher effort increases the probability of high performance, which boosts the reelection probability. This implies that voters can induce the incumbent's effort through reelection incentives as in the standard electoral accountability model (Besley, 2006).

Next, (ii) shows that $P^*(m,e)$ is increasing in the average confirmation bias among the political support base. As shown, voters among the political support base may misperceive low performance as high performance, and a higher confirmation bias increases this probability. Therefore, the higher the average confirmation bias is among the political support base, the less likely the incumbent will be punished. Consequently, the higher confirmation bias among the political support base is better for the incumbent. This point is in line with the finding by Lockwood (2017).

Lastly, as in (iii), the marginal benefit of effort, $\frac{\partial P^*(m,e)}{\partial e}$, is decreasing in the average confirmation bias among the political support base. The higher the average confirmation bias among the political support base, the less likely low performance is to be punished. Hence, the marginal benefit of exerting effort is reduced.

## 3.4 Average Confirmation Bias in the Political Support Base

As Lemma 3 (ii) indicates, a key determinant of the reelection probability is the average confirmation bias among the political support base; that is, $\mathbb{E}[q_i^t \mid I]$ when supporting the truth and $\mathbb{E}[q_i^t \mid M]$ when denying the truth. In this subsection, we explicitly derive these values because they will play a key role in the following analysis.

A simple calculation yields the explicit formula for these values:

$$\mathbb{E}[q_i^t \mid I] = \lambda \frac{\kappa \mathbb{E}[q] + (1-\kappa)[\mathbb{E}[q] - \mathbb{E}[q^2]]}{\kappa + (1-\kappa)[1-\mathbb{E}[q]]} + (1-\lambda)\mathbb{E}[q]; \tag{6}$$

$$\mathbb{E}[q_i^t \mid M] = \lambda \frac{\mathbb{E}[q^2]}{\mathbb{E}[q]} + (1-\lambda)\mathbb{E}[q]. \tag{7}$$

By comparing these two values, we have the following result:

**Lemma 4.** *$\mathbb{E}[q_i^t \mid I] \leq \mathbb{E}[q_i^t \mid M]$, where the equality holds if and only if $\sigma^2 = 0$. Furthermore, $\mathbb{E}[q_i^t \mid I]$ is decreasing in $\sigma^2$, whereas $\mathbb{E}[q_i^t \mid M]$ is increasing in $\sigma^2$.*

Suppose first that the degree of confirmation bias is homogeneous across voters (i.e., $\sigma^2 = 0$). By definition, in this case, the degree of confirmation bias is the same between informed and misinformed voters. On the contrary, the above lemma indicates that the average degree of confirmation bias regarding the incumbent's type among informed voters is strictly lower than that among misinformed voters when bias is heterogeneous (i.e., $\sigma^2 > 0$). That is, denying the truth means that voters with high confirmation bias become the political support base, which will be key in understanding our main result. Furthermore, this difference in the average degree of bias is increasing in $\sigma^2$.

The intuition for the case of heterogeneous bias can be understood as follows. In our model, voters receive a public signal containing the truth at the beginning of the game. Hence, if voters are Bayesian rational, all of them should become informed. However, misinformed voters exist because a part of voters may misperceive the public signal due to confirmation bias. Since the probability of misperception is increasing in the degree of confirmation bias, this implies that the average bias for the controversial issue is higher among misinformed voters. Furthermore, the degree of confirmation bias on the incumbent's type is correlated with that on the controversial issue. Therefore, we obtain the above result.

### 3.5 When the Incumbent Denies the Truth

Having these results in hand, we are now ready to present our main result.

**Case with low heterogeneity:** We first consider the case where the degree of confirmation bias is relatively homogeneous across voters i.e., the value of $\sigma^2$ is not so high. In this case, it is shown that the incumbent never denies the truth except for the case where she is the crazy type who non-strategically denies the truth.

**Proposition 1.** *Suppose that $\sigma^2 \leq \bar{\sigma}^2 := \frac{\mathbb{E}[q][1-2(1-\kappa)\mathbb{E}[q]]}{2\lambda(1-\kappa)}$. Then, in equilibrium, for any $\delta$, the opportunistic-type incumbent supports the truth (i.e., $m^*(\delta, 1) = 1$) as long as she is strategic.*

This result can be understood in a straightforward manner. To see this, suppose an extreme case where $\sigma^2 = 0$. As Lemma 4 indicates, $\sigma^2 = 0$ implies $\mathbb{E}[q_i^t \mid I] = \mathbb{E}[q_i^t \mid M]$; thus, from Lemma 3, $P^*(1, e) > P^*(0, e) \Leftrightarrow \kappa + (1-\kappa)(1-\mathbb{E}[q]) > (1-\kappa)\mathbb{E}[q]$, which is equivalent to Assumption 1. Hence, $P^*(1, e) > P^*(0, e)$ holds, meaning that supporting the truth boosts the reelection probability. Therefore, it is always optimal to support the truth.

In words, denying the truth shrinks the political support base. Since the smaller political support base reduces the number of obtained votes, the incumbent should not deny the truth . As such, the strategic incumbent never denies the truth as long as a majority of voters are informed. This result remains as long as the heterogeneity of confirmation bias is not so large.

**Case with high heterogeneity:** However, this no longer holds when the degree of confirmation bias is sufficiently heterogeneous across voters, which is summarized in the following proposition.

**Proposition 2.** *Suppose that $\sigma^2 > \bar{\sigma}^2$.*

*(i). The equilibrium is characterized by a unique $\bar{\delta} \in (0, \Delta]$. In particular, as long as she is strategic, the opportunistic-type incumbent supports the truth (i.e., $m^*(\delta, 1) = 1$) if and only if $\delta \geq \bar{\delta}$.*

*(ii). Furthermore, $\bar{\delta}$ is increasing in $\sigma^2$.*

That is, when the degree of confirmation bias is sufficiently heterogeneous across voters, the opportunistic type may strategically deny the truth. In particular, the opportunistic type strategically does so if and only if her competence level is below a threshold $\bar{\delta}$. Furthermore, the threshold value for the competence level, $\bar{\delta}$, is increasing in the heterogeneity of confirmation bias. That is, the more heterogeneous confirmation bias is, the more likely the incumbent is to deny the truth. This paradoxical incentive arises despite that the majority of voters are informed and do not believe the false argument.

This result can be understood based on the properties derived in the previous subsections. The key is that the distribution of confirmation bias differs between informed and misinformed voters. In the presence of large heterogeneity of confirmation bias, the average degree of confirmation bias among informed voters is much lower than that among misinformed voters; that is, $\mathbb{E}[q_i^t \mid M] > \mathbb{E}[q_i^t \mid I]$ holds (see Lemma 4). As a result, targeting misinformed voters as a political support base enables the incumbent to weaken the electoral punishment after the low performance (Lemma 3 (ii)-(iii)).

Therefore, the incumbent faces both pros and cons of denying the truth. On the one hand, its cost is that her political support base shrinks. On the other hand, its benefit is that her political support base has a higher confirmation bias so that electoral punishment for low performance becomes weaker. The performance of low-competent politicians are expected to be low; thus, for them, the benefit dominates the cost when the degree of confirmation bias is sufficiently heterogeneous across voters. As such, low-competent politicians take a seemingly paradoxical behavior: deny the truth supported by a majority of voters in order to capture voters with high confirmation bias as the political support base.

**Comparative statics:** Before concluding this subsection, we summarize the comparative statics result regarding how the value of $\bar{\sigma}^2$ is determined.

**Corollary 1.** *$\bar{\sigma}^2$ is decreasing in $\lambda$, but increasing in $\kappa$.[18] In addition, it is increasing in $\mathbb{E}[q]$ if and only if $\kappa > 1 - \frac{1}{4\mathbb{E}[q]}$.*

The first determinant of $\bar{\sigma}^2$ is the correlation of confirmation bias across issues, $\lambda$. The above corollary shows that the larger correlation induces the incumbent to deny the truth. To see the mechanism, suppose an extreme case where the correlation is zero (i.e., $\lambda = 0$). Then, voters believing the false argument have a high confirmation bias when learning the truth on this issue, but their confirmation bias when updating the incumbent's type is not necessarily low. Hence, attracting voters believing the false argument is not helpful in weakening electoral punishment. Therefore, when the correlation is low, the incumbent has no incentive to deny the truth. Though the mechanism is quite straightforward, its implication is significant. This result indicates that the effect of heterogeneity of confirmation bias depends on its type. When it is

[18]The upper bound of $\sigma^2$ is $\mathbb{E}[q](1 - \mathbb{E}[q])$. When either $\lambda = 0$ or $\kappa \leq 1/2$, $\bar{\sigma}^2 \geq \mathbb{E}[q](1 - \mathbb{E}[q])$ holds so that there exists no $\sigma^2$ such that $\bar{\delta} > 0$.

about the correlation across voters, the higher heterogeneity induces strategic misinformation, but when it is about the correlation across issues, the opposite is the case.

The second determinant is the fraction of voters who initially believed the truth, $\kappa$. The larger $\kappa$ means that the larger number of voters support the truth. Hence, it lowers the incumbent's incentive in denying the truth.

The last determinant is the average degree of confirmation bias, $\mathbb{E}[q]$. The higher confirmation bias has three effects. First, it increases the average degree of bias among informed voters. Second, the same holds among misinformed voters, too. Third, it increases the number of misinformed voters. The first one lowers the incumbent's incentive in denying the truth, whereas the second and third ones have the opposite effect. Hence, which dominates the other determines the total effect. On the one hand, when $\kappa$ is high, the fraction of informed voters is large so that the first effect dominates the second and third effects, implying that higher $\mathbb{E}[q]$ discourages the incumbent's misinformation. On the other hand, when $\kappa$ is low, the fraction of misinformed voters is large; thus higher $\mathbb{E}[q]$ encourages the incumbent's misinformation.

## 3.6 Effect on the Incumbent's Effort for Public Goods Provision

We have shown that denying the truth weakens electoral punishment. Hence, denying the truth might influence the incumbent's effort level for public goods provision. To examine this possibility, let the equilibrium effort level of the opportunistic type if she is strategic be $e^{**}(\delta) := e^*(\delta, m^*(\delta, 1))$.

Regarding this issue, we obtain the following proposition about the incumbent's effort level.

**Proposition 3.** *(i). Suppose that $\sigma^2 \leq \bar{\sigma}^2$. Then,*

$$e^{**}(\delta) = \frac{\delta}{(1-\varepsilon)a\Delta}\left(-1 + \sqrt{1 + 2(1-\varepsilon)a^2 b\psi\Delta[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}\right).$$

*Furthermore, $e^{**}(\delta)$ is increasing in $\delta$.*

*(ii). Suppose that $\sigma^2 > \bar{\sigma}^2$. If $\delta \geq \bar{\delta}$,*

$$e^{**}(\delta) = \frac{\delta}{(1-\varepsilon)a(\Delta^2 - \bar{\delta}^2)} \left(-\Delta + \sqrt{\Delta^2 + 2(1-\varepsilon)a^2 b\psi(\Delta^2 - \bar{\delta}^2)[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}\right),$$

*whereas if $\delta < \bar{\delta}$,*

$$e^{**}(\delta) = \frac{2\delta ab\psi(1-\kappa)\mathbb{E}[q](1-\mathbb{E}[q_i^t \mid M])\Delta}{\Delta + \sqrt{\Delta^2 + 2(1-\varepsilon)a^2 b\psi(\Delta^2 - \bar{\delta}^2)[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}}.$$

*Furthermore, $e^{**}(\delta)$ is increasing in $\delta$ and $\lim_{\delta \nearrow \bar{\delta}} e^{**}(\delta) < \lim_{\delta \searrow \bar{\delta}} e^{**}(\delta)$.*

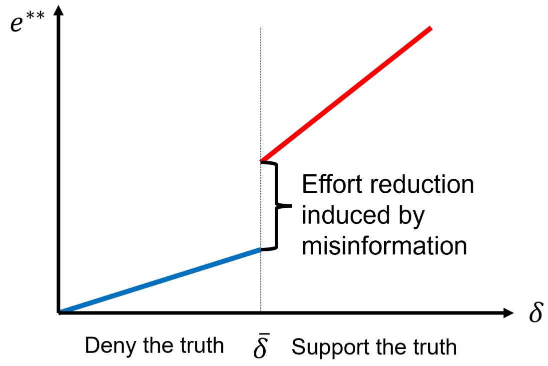


Figure 2: Effort, Misinformation, and Competence

First, suppose that the extent of heterogeneity of confirmation bias is less than $\bar{\sigma}^2$ so that the opportunistic type never denies the truth as long as she is strategic. Then, the effort level is given as in (i) of the proposition and it is increasing in $\delta$ because the higher $\delta$ means the lower marginal cost of effort.

Second, suppose that the extent of heterogeneity of confirmation bias exceeds $\bar{\sigma}^2$; that is, the opportunistic type never denies the truth if and only if $\delta \geq \bar{\delta}$. In this case, the level of effort depends on whether $\delta \geq \bar{\delta}$ as in (ii) of the proposition, which is also visualized in Figure 2. A distinctive feature of this case is that the effort level discontinuously drops when $\delta$ becomes below the threshold, despite that the cost of effort is close to each other around the threshold. This discontinuous drop identifies the negative effect of strategic misinformation. When the incumbent denies the truth, her political support base is misinformed voters who are less likely to punish low performance. As a result, the marginal benefit of effort discontinuously drops once the incumbent denies the truth, which reduces the effort level.

This finding reveals the hidden welfare cost of misinformation. It would be believed that politicians' denying the truth induces ordinary citizens to hold wrong beliefs. For example, the president Bolsonaro's misinformation about COVID-19 in Brazil is criticized because it forces the Brazilians to face severe COVID-19 situations (Ricard and Medeiros, 2020). While we share the same concern, our result uncovers another hidden cost of misinformation. As shown above, low-competent politicians deny the truth to weaken the electoral punishment for low performance. Hence, misinformation has a negative spill-over effect: it lowers politicians' effort level for providing public goods, which reduces voters' welfare. This hidden cost should be considered in evaluating the welfare cost of misinformation.

**Effect of heterogeneous confirmation bias:** As shown above, denying the truth reduces the effort level. Furthermore, the opportunistic type denies the truth if and only if the extent of heterogeneity of confirmation bias is sufficiently large. Given these findings, it might be expected that the larger heterogeneity of confirmation bias reduces the effort level. Interestingly, this naive view is not necessarily true. The effect of heterogeneous confirmation bias could be non-monotonic. To see this, let $e^{**}(\delta)$ given $\sigma^2$ be $e^{**}(\delta \mid \sigma^2)$ and $\bar{\delta}$ given $\sigma^2$ be $\bar{\delta}(\sigma^2)$.

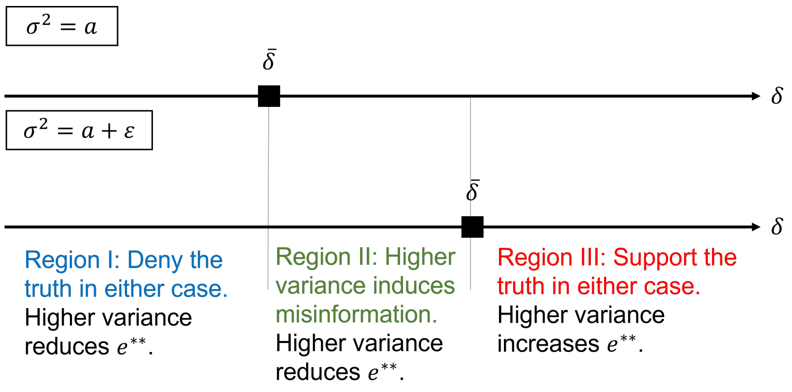


Figure 3: Larger Heterogeneity and Effort

**Proposition 4.** *Suppose that $\sigma^2$ changes from $a$ to $a + \varepsilon$, where $\varepsilon$ is a sufficiently small, but positive value.*

*(i). Suppose that $a < \bar{\sigma}^2$. Then, $e^{**}(\delta \mid a) < e^{**}(\delta \mid a + \varepsilon)$ holds for any $\delta$.*

*(ii). Suppose that $a > \bar{\sigma}^2$. Then, $e^{**}(\delta \mid a) > e^{**}(\delta \mid a + \varepsilon)$ holds for any $\delta \in (0, \bar{\delta}(a + \varepsilon))$, but $e^{**}(\delta \mid a) < e^{**}(\delta \mid a + \varepsilon)$ holds for any $\delta \in [\bar{\delta}(a + \varepsilon), \Delta)$.*

First, suppose that the initial value of $\sigma^2$ is small. Then, as in (i) of the above proposition, a marginal change of $\sigma^2$ not reduces but increases the opportunistic type's effort level; thus, the larger heterogeneity is beneficial for voters when the level of heterogeneity is initially low. The mechanism is understood as follows. As the larger heterogeneity reduces the average confirmation bias among informed voters (Lemma 4), it stimulates the effort level of politicians whose political support base is informed voters. Given the low level of heterogeneity, every politician supports the truth so that her political support base is always informed voters. Therefore, a marginal increase in $\sigma^2$ has a positive effect on every politician's effort level. This result rejects the naive view inferred based on Proposition 2 that the larger variance is harmful because it induces strategic misinformation.

Second, suppose that the initial value of $\sigma^2$ is large. Then, as in (ii) of the above proposition, a marginal change of $\sigma^2$ reduces the opportunistic type's effort level if and only if $\delta$ falls below a threshold value. That is, whether the larger heterogeneity is harmful to voters depends on the incumbent's competence level. The mechanism is understood as follows. Consider the low-competent politicians who deny the truth in either value of $\sigma^2$, which corresponds to Region I in Figure 3. The greater heterogeneity increases the average confirmation bias among their political support base, misinformed voters (Lemma 4). Hence, the larger heterogeneity discourages effort. Next, consider politicians with an intermediate level of competence, which corresponds to Region II in Figure 3. They initially support the truth, but start to deny the truth after an increase in $\sigma^2$, which reduces the level of effort. Lastly, consider highly competent politicians, which corresponds to Region III in Figure 3. They support the truth in either case so that their political support base is informed voters. As the larger heterogeneity reduces the average confirmation bias among informed voters (Lemma 4), the effort level is stimulated for them, which is opposite to the effects on Regions I and II.

In sum, the effect of the larger heterogeneity on the effort level is non-monotonic as a result of the interaction with strategic misinformation. The larger heterogeneity induces low-competent politicians to strategically deny the truth and reduce their effort. However, it also encourages high-competent politicians' incentive to exert effort. This point should be taken into account when evaluating the welfare effect of the larger heterogeneity with respect to confirmation bias.

# 4 Discussion

## 4.1 Extensions

**Challenger's message:** Thus far, we have assumed that only the incumbent sends a cheap-talk message, $m$. However, the challenger may also decide whether to support or deny the truth. To examine this possibility, suppose that the incumbent and the challenger simultaneously send a cheap-talk message at the beginning of the game. The incumbent's (resp. the challenger's) message is denoted by $m_I$ (resp. $m_C$). Other settings remain the same. Under this setting, we obtain the following result:

**Proposition 5.** *On the one hand, the incumbent's strategy is the same as in Propositions 1 and 2. On the other hand, the opportunistic-type challenger never denies the truth as long as she is strategic (i.e., $m^*_C(\delta, 1) = 1$ for all $\delta$).*

Therefore, the challenger has no incentive to deny the truth, whereas the incumbent may do so in the presence of heterogeneous confirmation bias. The mechanism can be understood in a straightforward way. As we have shown in the main analysis, the incumbent denies the truth in order to weaken electoral punishment for the low performance in public goods provision. On the contrary, the challenger exerts no effort in public goods provision so that she is not punished by the incumbent's low performance. Hence, for the challenger, the average degree of confirmation bias among her political support base does not matter. What matters is only the size of her political support base. Therefore, denying the truth supported by a majority of voters is not optimal for the challenger.

Having said that, it should be noted that this proposition does not imply that the challenger is always less likely to deny the truth, because there exists a selection issue omitted from our model. In practice, some challengers may believe the false argument and deny the truth for non-strategic reasons, whereas it is unlikely to be the case for the incumbent because such politicians lose an election and cannot be the incumbent. In other words, the value of $\varepsilon$, the fraction of the non-strategic opportunistic type, is expected to be much larger for the challenger than for the incumbent. Since the probability of denying the truth is given by $\varepsilon + (1 - \varepsilon)\mathbb{E}[1 - m^*(\delta, 1)]$, the challenger may be more likely to deny the truth when this selection effect dominates the incentive effect presented in the above proposition. Which dominates the other will be left for future empirical research.

**Effect of social media:** Our results also provide an implication regarding the echo-chambers effects in recent information environments. As social media platforms have been rapidly developed, there is a widely shared concern that social media may create homogeneous and polarized groups (i.e., echo chambers) by allowing users to selectively expose themselves to information that reinforces their existing views (e.g., Del Vicario et al., 2016; Acemoglu, Ozdaglar and Siderius, 2021).

To examine how this echo-chamber effect influences the political supply of misinformation, it should be noted that, in practice, confirmation bias is a joint product of the psychological factor and the information environment, though we have emphasized the former factor. As mentioned, social media platforms enable people to selectively expose themselves to pro-attitudinal news, which amplifies confirmation bias stemming from psychological factors. Provided these points, voter $i$'s degree of confirmation bias, $q_i$, can be decomposed into two terms: $q_i = \tau q_{ib}$ where $\tau > 0$ is the feasibility of selective exposure determined by the information environment and $q_{ib}$ is voter $i$'s psychological bias. For simplicity, suppose that $q_{ib}$ follows $U[0, B]$; thus, $q_i$ follows $U[0, \tau B]$. Note that $\tau < 1/B$ is assumed.

The emergence of social media platforms increases the value of $\tau$, which has two effects on confirmation bias. First, the average confirmation bias, $\mathbb{E}[q_i]$, increases so that the number of misinformed voters increases. Second, the variance of $q_i$, $\sigma^2$, also increases because the variance is given by $\frac{\tau^2 B^2}{12}$. Hence, the bias becomes more heterogeneous across voters.

The first effect through an increase in $\mathbb{E}[q_i]$ corresponds to a widely shared concern about echo chambers (Del Vicario et al., 2016). As in Lemma 1, larger $\mathbb{E}[q_i]$ increases the number of misinformed voters. Hence, the first effect indicates that social media platforms make a large fraction of voters misinformed. Note that its effect on the political supply of misinformation is ambiguous as shown in Corollary 1: it fosters the incumbent to deny the truth if and only if $\kappa$ is sufficiently low.

In addition to this, our model identifies a novel effect through an increase in $\sigma^2$. Proposition 2 indicates that larger $\sigma^2$ fosters the political supply of misinformation. Therefore, the emergence of social media platforms encourages politicians to deny the truth not only through the higher average confirmation bias but also through larger heterogeneity of this bias between voters. The latter effect has been ignored in the literature, but our results suggest that the effect of social media platforms through larger heterogeneity should be considered in evaluation its effect on misinformation.

The total effect of higher $\tau$ can be obtained as the following proposition:

**Proposition 6.**

$$\sigma^2 > \bar{\sigma}^2 \;\Leftrightarrow\; \tau > \frac{3}{(\lambda+3)(1-\kappa)B}.$$

Hence, the incumbent denies the truth if and only if the echo-chamber effect of social media platforms is sufficiently large. As discussed above, this should not be seen as a result of the larger average confirmation bias. Depending on the value of $\kappa$, the larger average confirmation bias may discourage the incumbent to deny the truth. This proposition shows that the effect through the larger variance is a dominant force even in such a case. Consequently, the larger echo-chamber effect of social media platforms always encourages the incumbent to deny the truth.

## 4.2 Implications for Empirical Research

**Demand side of misinformation:** The literature on political psychology has devoted large effort to experimental research for understanding the role of psychological biases in accepting misinformation (Pantazi, Hale and Klein, 2021; Flynn, Nyhan and Reifler, 2017). However, except for several attempts (e.g., Taber and Lodge, 2006; Kahan et al., 2017; Arceneaux and Vander Wielen, 2017), most experimental studies of misinformation do not account for the heterogeneity of psychological biases. For example, as a future research agenda, Flynn, Nyhan and Reifler (2017) point out that "it would be helpful to construct and validate a scale (or scales) that measures individual-level differences in the strength of underlying accuracy and/or directional motivations" (p.137).

Our result indicates that the purpose of such attempts should not be limited to understanding individual differences in accepting misinformation. As Proposition 2 shows, we cannot evaluate the effect on politicians' incentive for misinformation and democratic accountability without identifying the extent of heterogeneity across voters. To empirically evaluate the effect on politics, we need to adopt the experimental design enabling us to identify individual-level psychological biases and derive their variance.

**Electoral (dis)advantage of misinformation:** Whether denying the truth fosters the chance of electoral winning or not is an important question (Barrera et al., 2020). If it is costly in winning an election, the spread of misinformation among politicians may not be so problematic for democracy because politicians denying the truth would eventually exit from the political arena.

Our results suggest that the following two issues should be handled in empirically evaluating the electoral (dis)advantage of misinformation. The first one is for empirical research based on observational data. A naive way to test the electoral (dis)advantage is to regress a politician's vote share on whether the politician supports misinformation. Our results show that this strategy suffers from the following endogeneity problem. As shown in Proposition 2, low-competent politicians are likely to support misinformation, but their performance tends to be low. Hence, without accounting for this, we may obtain a negative correlation between denying the truth and the electoral outcome even if denying the truth is indeed beneficial as in our model.

Furthermore, finding a (quasi-)experimental situation in which whether to deny the truth is determined in a random way is still unsatisfactory. The causal effect of denying the truth given $\delta$ is $P^*(1, e^*(\delta, 1)) - P^*(0, e^*(\delta, 0))$. This is heterogeneous depending on the value of competence level, $\delta$, because the incumbent's expected performance depends on it. In particular, it can be easily shown that it is positive for small $\delta$, but negative for large $\delta$ when $\sigma^2 > \bar{\sigma}^2$. That is, even the direction of the causal effect depends on the incumbent's level of competence. Hence, quantifying each politician's level of competence is essential.

The second issue is in conducting experimental research. A straightforward approach would be to conduct a conjoint survey experiment where each subject is assigned to a hypothetical politician with various attributes and asked whether to support the politician or not (Bansak et al., 2021). Attributes should include whether the politician supports misinformation. Then, we could identify the electoral effect of denying the truth. However, this approach should be improved. Our model indicates that the electoral benefit of denying the truth is stemming from the fact that it weakens the electoral punishment when low performance is realized. Hence, it is not appropriate to ask voters' support for a politician just after the politician denies the truth. What we need to measure is voters' support for the politician when her low performance is realized sometime after she denied the truth. Conjoint survey experiments should be designed to accommodate this issue.[19]

# 5 Concluding Remarks

The current era is characterized by the spread of misinformation, fake news, and conspiracy theories that are evidently false. A notable feature of this spread is that even elected politicians sometimes support false claims. We analyzed this political supply of misinformation from the perspective of politicians' electoral incentives. Specifically, we constructed a two-period electoral accountability model where the incumbent politician decides whether to support or deny the truth about a particular issue at the beginning of the game. We showed that the low-competent incumbent paradoxically has an electoral incentive to supply misinformation even if only a minority of voters believe the misinformation. The key to understanding this paradox is the heterogeneity of confirmation bias, a psychological bias that has been widely regarded as a determinant of people's persistent acceptance of misinformation. We showed that the low-competent politician denies the truth if and only if the degree of confirmation bias is sufficiently heterogeneous across voters. This result indicates that individual differences in confirmation bias play a key role in the political supply of misinformation.

# References

**Acemoglu, Daron, Asuman E Ozdaglar, and James Siderius.** 2021. "A model of online misinformation." *Working Paper*.

**Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin.** 2013. "A political theory of populism." *The Quarterly Journal of Economics*, 128(2): 771–805.

[19]Including the politician's performance as an attribute in an experiment does not resolve this issue because low performance and denying the truth are presented at the same time.

**Andreoni, James, and Tymofiy Mylovanov.** 2012. "Diverging opinions." *American Economic Journal: Microeconomics*, 4(1): 209–32.

**Arceneaux, Kevin, and Ryan J Vander Wielen.** 2017. *Taming intuition: How reflection minimizes partisan reasoning and promotes democratic accountability.* Cambridge University Press.

**Bansak, Kirk, Jens Hainmueller, Daniel J Hopkins, and Teppei Yamamoto.** 2021. "Conjoint survey experiments." *Cambridge Handbook of Advances in Experimental Political Science*, 19–41.

**Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya.** 2020. "Facts, alternative facts, and fact checking in times of post-truth politics." *Journal of Public Economics*, 182: 104123.

**Bénabou, Roland, and Jean Tirole.** 2016. "Mindful economics: The production, consumption, and value of beliefs." *Journal of Economic Perspectives*, 30(3): 141–64.

**Besley, Timothy.** 2006. *Principled agents?: The political economy of good government.* Oxford University Press.

**Carrillo, Juan D, and Micael Castanheira.** 2008. "Information and strategic political polarisation." *The Economic Journal*, 118(530): 845–874.

**Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi.** 2016. "The spreading of misinformation online." *Proceedings of the National Academy of Sciences*, 113(3): 554–559.

**Dixit, Avinash, and John Londregan.** 1996. "The determinants of success of special interests in redistributive politics." *The Journal of Politics*, 58(4): 1132–1155.

**Duggan, John, and César Martinelli.** 2017. "The political economy of dynamic elections: Accountability, commitment, and responsiveness." *Journal of Economic Literature*, 55(3): 916–84.

**Eguia, Jon X, and Francesco Giovannoni.** 2019. "Tactical extremism." *American Political Science Review*, 113(1): 282–286.

**Enke, Benjamin, and Florian Zimmermann.** 2019. "Correlation neglect in belief formation." *The Review of Economic Studies*, 86(1): 313–332.

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

**Fryer, Roland G Jr, Philipp Harms, and Matthew O Jackson.** 2019. "Updating beliefs when evidence is open to interpretation: Implications for bias and polarization." *Journal of the European Economic Association*, 17(5): 1470–1501.

**Gehlbach, Scott, and Konstantin Sonin.** 2014. "Government control of the media." *Journal of Public Economics*, 118: 163–171.

**Glaeser, Edward L.** 2005. "The political economy of hatred." *The Quarterly Journal of Economics*, 120(1): 45–86.

**Glaeser, Edward L, Giacomo AM Ponzetto, and Jesse M Shapiro.** 2005. "Strategic extremism: Why Republicans and Democrats divide on religious values." *The Quarterly Journal of Economics*, 120(4): 1283–1330.

**Grillo, Edoardo, and Carlo Prato.** forthcoming. "Reference points and democratic backsliding." *American Journal of Political Science*.

**Grossman, Gene M, and Elhanan Helpman.** 2022. "Electoral competition with fake news." *European Journal of Political Economy*, 102315.

**Guriev, Sergei, and Daniel Treisman.** 2020. "A theory of informational autocracy." *Journal of Public Economics*, 186: 104158.

**Horz, Carlo M.** 2021. "Propaganda and skepticism." *American Journal of Political Science*, 65(3): 717–732.

**Jerit, Jennifer, and Yangzi Zhao.** 2020. "Political misinformation." *Annual Review of Political Science*, 23(1): 77–94.

**Kahan, Dan M, Asheley Landrum, Katie Carpenter, Laura Helft, and Kathleen Hall Jamieson.** 2017. "Science curiosity and political information processing." *Political Psychology*, 38: 179–199.

**Kartal, Melis, and Jean-Robert Tyran.** 2022. "Fake news, voter overconfidence, and the quality of democratic choice." *American Economic Review*, 112(10): 3367–97.

**Lasser, Jana, Segun Taofeek Aroyehun, Almog Simchon, Fabio Carrella, David Garcia, and Stephan Lewandowsky.** forthcoming. "Social media sharing of low quality news sources by political elites." *PNAS Nexus*.

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

**Le Yaouanq, Yves.** 2021. "Motivated cognition in a model of voting." *Available at SSRN 3974501*.

**Lindbeck, Assar, and Jörgen W Weibull.** 1987. "Balanced-budget redistribution as the outcome of political competition." *Public Choice*, 52(3): 273–297.

**Little, Andrew T.** 2017. "Propaganda and credulity." *Games and Economic Behavior*, 102: 224–232.

**Little, Andrew T.** 2022. "Detecting motivated reasoning." *Working Paper*.

**Little, Andrew T, Keith E Schnakenberg, and Ian R Turner.** 2022. "Motivated reasoning and democratic accountability." *American Political Science Review*, 116(2): 751–767.

**Lockwood, Ben.** 2017. "Confirmation bias and electoral accountability." *Quarterly Journal of Political Science*, 11(4): 471–501.

**Loh, Isaac, and Gregory Phelan.** 2019. "Dimensionality and disagreement: Asymptotic belief divergence in response to common information." *International Economic Review*, 60(4): 1861–1876.

**Lord, Charles G, Lee Ross, and Mark R Lepper.** 1979. "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence." *Journal of Personality and Social Psychology*, 37(11): 2098.

**Millner, Antony, Hélène Ollivier, and Leo Simon.** 2020. "Confirmation bias and signaling in Downsian elections." *Journal of Public Economics*, 185: 104175.

**Nickerson, Raymond S.** 1998. "Confirmation bias: A ubiquitous phenomenon in many guises." *Review of General Psychology*, 2(2): 175–220.

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

**Nyhan, Brendan, and Jason Reifler.** 2010. "When corrections fail: The persistence of political misperceptions." *Political Behavior*, 32(2): 303–330.

**Nyhan, Brendan, and Jason Reifler.** 2015. "The effect of fact-checking on elites: A field experiment on US state legislators." *American Journal of Political Science*, 59(3): 628–640.

**Pantazi, Myrto, Scott Hale, and Olivier Klein.** 2021. "Social and cognitive aspects of the vulnerability to political misinformation." *Political Psychology*, 42: 267–304.

**Rabin, Matthew, and Joel L Schrag.** 1999. "First impressions matter: A model of confirmatory bias." *The Quarterly Journal of Economics*, 114(1): 37–82.

**Ricard, Julie, and Juliano Medeiros.** 2020. "Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil." *Harvard Kennedy School Misinformation Review*, 1(3).

**Swire-Thompson, Briony, Joseph DeGutis, and David Lazer.** 2020. "Searching for the backfire effect: Measurement and design considerations." *Journal of Applied Research in Memory and Cognition*, 9(3): 286–299.

**Swire-Thompson, Briony, Ullrich KH Ecker, Stephan Lewandowsky, and Adam J Berinsky.** 2020. "They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation." *Political Psychology*, 41(1): 21–34.

**Szeidl, Adam, and Ferenc Szucs.** 2022. "The political economy of alternative realities." *Working Paper*.

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

**Taber, Charles S, Damon Cann, and Simona Kucsova.** 2009. "The motivated processing of political arguments." *Political Behavior*, 31(2): 137–155.

**Uscinski, Joseph E, Adam M Enders, Casey Klofstad, Michelle Seelig, John Funchion, Caleb Everett, Stefan Wuchty, Kamal Premaratne, and Manohar Murthi.** 2020. "Why do people believe COVID-19 conspiracy theories?" *Harvard Kennedy School Misinformation Review*, 1(3).

**Wilson, Andrea.** 2014. "Bounded memory and biases in information processing." *Econometrica*, 82(6): 2257–2294.

**Yariv, Leeat.** 2002. "I'll see it when I believe it—A simple model of cognitive consistency." *Working Paper*.

# A Omitted Proofs

## A.1 Proof of Lemma 4

We first derive the values of $\mathbb{E}[q_i^t \mid I]$ and $\mathbb{E}[q_i^t \mid M]$.

$$\mathbb{E}[q_i^t \mid I] = \lambda\mathbb{E}[q_i^\omega \mid I] + (1-\lambda)\mathbb{E}[q],$$

where

$$\begin{aligned}
\mathbb{E}[q_i^\omega \mid I] =& \frac{\kappa}{\kappa + (1-\kappa)(1-\mathbb{E}[q])}\mathbb{E}[q] + \frac{(1-\kappa)(1-\mathbb{E}[q])}{\kappa + (1-\kappa)(1-\mathbb{E}[q])}\int_0^1 q_i^\omega \frac{(1-q_i^\omega)f(q_i^\omega)}{1-\mathbb{E}[q]} dq_i^\omega \\
=& \frac{\kappa}{\kappa + (1-\kappa)(1-\mathbb{E}[q])}\mathbb{E}[q] + \frac{1-\kappa}{\kappa + (1-\kappa)(1-\mathbb{E}[q])}\mathbb{E}[q(1-q)] \\
=& \frac{\kappa\mathbb{E}[q] + (1-\kappa)(\mathbb{E}[q] - \mathbb{E}[q^2])}{\kappa + (1-\kappa)(1-\mathbb{E}[q])}.
\end{aligned}$$

Note that $f$ is a density function.[20]

Similarly,

$$\mathbb{E}[q_i^t \mid M] = \lambda \mathbb{E}[q_i^\omega \mid M] + (1-\lambda)\mathbb{E}[q],$$

where

$$\mathbb{E}[q_i^\omega \mid M] = \int_0^1 q_i^\omega \frac{q_i^\omega f(q_i^\omega)}{\mathbb{E}[q]} dq_i^\omega = \frac{\mathbb{E}[q^2]}{\mathbb{E}[q]}.$$

We prove that $\mathbb{E}[q_i^t \mid I] \leq \mathbb{E}[q_i^t \mid M]$ holds given the derived values of $\mathbb{E}[q_i^t \mid I]$ and $\mathbb{E}[q_i^t \mid M]$.

$$\begin{aligned}
\mathbb{E}[q_i^t \mid I] \leq \mathbb{E}[q_i^t \mid M] &\Leftrightarrow \lambda \frac{\kappa \mathbb{E}[q] + (1-\kappa)[\mathbb{E}[q] - \mathbb{E}[q^2]]}{\kappa + (1-\kappa)[1-\mathbb{E}[q]]} + (1-\lambda)\mathbb{E}[q] \leq \lambda \frac{\mathbb{E}[q^2]}{\mathbb{E}[q]} + (1-\lambda)\mathbb{E}[q] \\
&\Leftrightarrow \mathbb{E}[q]^2 - (1-\kappa)\mathbb{E}[q^2]\mathbb{E}[q] \leq \mathbb{E}[q^2] - (1-\kappa)\mathbb{E}[q^2]\mathbb{E}[q] \\
&\Leftrightarrow \mathbb{E}[q]^2 \leq \mathbb{E}[q^2].
\end{aligned}$$

In general, $\mathbb{E}[q^2] - \mathbb{E}[q]^2 = \sigma^2 \geq 0$ holds. Hence, the above inequality directly implies the first part of the lemma.

Furthermore, by substituting $\mathbb{E}[q^2] = \mathbb{E}[q]^2 + \sigma^2$ into $\mathbb{E}[q_i^t \mid I]$ and $\mathbb{E}[q_i^t \mid M]$, we have the second part of the lemma. □

## A.2 Preliminary for the Proofs of Propositions 1 and 2

Before proving these propositions, we prove the following lemma, which states that the opportunistic type's communication strategy is characterized by a threshold property. Note that this result is applicable to the cases with any $\sigma^2$.

**Lemma 5.** *For any $\sigma^2$, the following holds. There exists $\bar{\delta} \in [0, \Delta]$ such that $m^*(\delta, 1) = 1$ (resp. $m^*(\delta, 1) = 0$) if $\delta > \bar{\delta}$ (resp. $\delta < \bar{\delta}$).*

*Proof.* First, we derive the optimal effort level given $m = 1$, which is the solution to the following maximization problem:

$$\max_e U(1, e) := P^*(1, e)b - \frac{e^2}{2\delta}.$$

By taking the first-order condition, we obtain

$$e^*(\delta, 1) = \delta a b \psi \frac{\kappa + (1-\kappa)(1-\mathbb{E}[q])}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} \bar{p}(1 - \mathbb{E}[q_i^t \mid I]).$$

[20]We allow the distribution to be discrete. In that case, we consider a discrete version of $\int_0^1 q_i^\omega \frac{(1-q_i^\omega) f(q_i^\omega)}{1-\mathbb{E}[q]} dq_i^\omega$; then, the same formula can be derived.

Substituting this into $U(1, e)$ yields the incumbent's value function when $m = 1$:

$$\begin{aligned} U(1, e^*) =& \frac{1-\psi}{2} b + \frac{\kappa + (1-\kappa)(1-\mathbb{E}[q])}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} b\psi \bar{p} \mathbb{E}[q_i^t \mid I] \\ &+ \frac{\delta a^2 b^2 \psi^2}{2} \left( \frac{\kappa + (1-\kappa)(1-\mathbb{E}[q])}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} \right)^2 \bar{p}^2 (1 - \mathbb{E}[q_i^t \mid I])^2. \end{aligned}$$

Next, we derive the optimal effort level given $m = 0$, which is the solution to the following maximization problem:

$$\max_e U(0, e) := P^*(0, e) - \frac{e^2}{2\delta}.$$

By taking the first-order condition, we obtain

$$e^*(\delta, 0) = \delta a b \psi \frac{(1-\kappa)\mathbb{E}[q]}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} \bar{p} (1 - \mathbb{E}[q_i^t \mid M]).$$

Substituting this into $U(0, e)$ yields the incumbent's value function when $m = 0$:

$$\begin{aligned} U(0, e^*) =& \frac{1-\psi}{2} b + \frac{(1-\kappa)\mathbb{E}[q]}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} b\psi \bar{p} \mathbb{E}[q_i^t \mid M] \\ &+ \frac{\delta a^2 b^2 \psi^2}{2} \left( \frac{(1-\kappa)\mathbb{E}[q]}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} \right)^2 \bar{p}^2 (1 - \mathbb{E}[q_i^t \mid M])^2. \end{aligned}$$

The incumbent has a strict incentive to send message 1 if and only if $U(1, e^*) > U(0, e^*)$. Hence, it suffices to prove that $U(1, e^*) - U(0, e^*)$ is increasing in $\delta$. Indeed,

$$\begin{aligned} &\frac{\partial [U(1, e^*) - U(0, e^*)]}{\partial \delta} > 0 \\ \Leftrightarrow & [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1 - \mathbb{E}[q_i^t \mid I]) > [(1-\kappa)\mathbb{E}[q]](1 - \mathbb{E}[q_i^t \mid M]), \end{aligned}$$

which holds because $\kappa + (1-\kappa)(1-\mathbb{E}[q]) > (1-\kappa)\mathbb{E}[q]$ from Assumption 1 and $\mathbb{E}[q_i^t \mid I] \leq \mathbb{E}[q_i^t \mid M]$ from Lemma 4. Therefore, we obtain the desired result. □

## A.3 Proof of Proposition 1

It suffices to prove that $\bar{\delta} = 0$; that is, the opportunistic type must support the truth as long as she is strategic. For this purpose, it suffices to prove that $U(1, e^*) \geq U(0, e^*)$ when $\delta = 0$. When

$\delta = 0$,

$$
\begin{aligned}
& U(1, e^*) \geq U(0, e^*) \\
\Leftrightarrow & [\kappa + (1-\kappa)(1-\mathbb{E}[q])]\mathbb{E}[q_i^t \mid I] \geq [(1-\kappa)\mathbb{E}[q]]\mathbb{E}[q_i^t \mid M] \\
\Leftrightarrow & \lambda[\kappa\mathbb{E}[q] + (1-\kappa)(\mathbb{E}[q] - \mathbb{E}[q]^2 - \sigma^2)] + (1-\lambda)\kappa\mathbb{E}[q] + (1-\lambda)(1-\kappa)(\mathbb{E}[q] - \mathbb{E}[q]^2) \\
& \quad \geq \lambda(1-\kappa)(\mathbb{E}[q]^2 + \sigma^2) + (1-\lambda)(1-\kappa)\mathbb{E}[q]^2 \\
\Leftrightarrow & \sigma^2 \leq \bar{\sigma}^2 = \frac{\mathbb{E}[q]\,[1 - 2(1-\kappa)\mathbb{E}[q]]}{2\lambda(1-\kappa)}
\end{aligned}
$$

where the third line is obtained by substituting (6), (7), and $\mathbb{E}[q^2] = \mathbb{E}[q]^2 + \sigma^2$ into the inequality. Therefore, under our assumption, $U(1, e^*) \geq U(0, e^*)$ when $\delta = 0$, implying that $\bar{\delta} = 0$. □

## A.4 Proof of Proposition 2 (i)

$P^*(m, e)$ depends on $\bar{\delta}$ and $e^*(\delta, m)$ also depends on $\bar{\delta}$. To make this point clear, we denote $P^*(m, e)$ by $P^*(m, e \mid \bar{\delta})$ and $e^*(\delta, m)$ by $e^*(\delta, m \mid \bar{\delta})$ respectively. Then, the opportunistic type with $\delta$ has an incentive to tell the truth if and only if

$$U(1, e^*(\delta, 1 \mid \bar{\delta})) \geq U(0, e^*(\delta, 0 \mid \bar{\delta})).$$

Specifically, if $\bar{\delta}$ is an interior (i.e., $\bar{\delta} \in (0, \Delta)$),

$$U(1, e^*(\bar{\delta}, 1 \mid \bar{\delta})) = U(0, e^*(\bar{\delta}, 0 \mid \bar{\delta}))$$

holds. To examine this condition, let

$$V(\delta) := U(1, e^*(\delta, 1 \mid \delta)) - U(0, e^*(\delta, 0 \mid \delta)).$$

Since $\sigma^2 > \bar{\sigma}^2$, $V(0) < 0$ holds from the proof of Proposition 1. Hence, there is no equilibrium in which $\bar{\delta} = 0$.

Step 1. As a first step, we show that there exists at most one $\delta$ such that $V(\delta) = 0$ is satisfied. For this purpose, it suffices to prove that $V'(\delta) > 0$ when $V(\delta) \geq 0$. Let $P^*(\delta, m \mid \bar{\delta})$ be $P^*(m, e^*(\delta, m) \mid \bar{\delta})$. A simple calculation yields:

$$
\begin{aligned}
V'(\delta) =& b\left(\frac{\partial P^*}{\partial \delta}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \delta}(\delta, 0 \mid \delta)\right) + \frac{1}{2\delta^2}\left(e^*(\delta, 1 \mid \delta))^2 - e^*(\delta, 0 \mid \delta))^2\right) \\
& + \frac{\partial e^*(\delta, 1 \mid \delta)}{\partial \delta}\frac{\partial U}{\partial e}(1, e^*(\delta, 1 \mid \delta)) - \frac{\partial e^*(\delta, 0 \mid \delta)}{\partial \delta}\frac{\partial U}{\partial e}(1, e^*(\delta, 0 \mid \delta)) \\
=& b\left(\frac{\partial P^*}{\partial \delta}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \delta}(\delta, 0 \mid \delta)\right) + \frac{1}{2\delta^2}\left(e^*(\delta, 1 \mid \delta))^2 - e^*(\delta, 0 \mid \delta))^2\right),
\end{aligned}
$$

where the second equality comes from the fact that

$$\frac{\partial U}{\partial e}(1, e^*(\delta, 1 \mid \delta)) = \frac{\partial U}{\partial e}(0, e^*(\delta, 0 \mid \delta)) = 0$$

because of the envelope theorem.

(i). $\frac{\partial P^*}{\partial \delta}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \delta}(\delta, 0 \mid \delta) \geq 0$.

To show this, by applying the definition of $\bar{p}$, we have

$$\frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} = \frac{1}{1 + (1-\varepsilon)a\int_{\bar{\delta}}^{\Delta} \frac{e^*(1,\tilde{\delta})}{\Delta} d\tilde{\delta}}.$$

Hence,

$$\begin{aligned}
&\frac{\partial P^*}{\partial \delta}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \delta}(\delta, 0 \mid \delta) \\
&= \psi \left\{ [\kappa + (1-\kappa)(1-\mathbb{E}[q])] \left[ae^*(\delta, 1 \mid \delta) + \mathbb{E}[q_i^t \mid I](1 - ae^*(\delta, 1 \mid \delta))\right] \right. \\
&\quad \left. -(1-\kappa)\mathbb{E}[q] \left[ae^*(\delta, 0 \mid \delta) + \mathbb{E}[q_i^t \mid M](1 - ae^*(\delta, 0 \mid \delta))\right] \right\} \\
&\quad \times \frac{\partial}{\partial \delta}\left(\frac{1}{1 + (1-\varepsilon)a\int_{\delta}^{\Delta} \frac{e^*(1,\tilde{\delta})}{\Delta} d\tilde{\delta}}\right).
\end{aligned} \tag{8}$$

Here, the last term is obviously positive. In addition, regarding the first term,

$$\begin{aligned}
V(\delta) \geq 0 \Leftrightarrow & P^*(\delta, 1 \mid \delta) \geq P^*(\delta, 0 \mid \delta) \\
& \Leftrightarrow [\kappa + (1-\kappa)(1-\mathbb{E}[q])] \left[ae^*(\delta, 1 \mid \delta) + \mathbb{E}[q_i^t \mid I](1 - ae^*(\delta, 1 \mid \delta))\right] \\
& \quad \geq (1-\kappa)\mathbb{E}[q] \left[ae^*(\delta, 0 \mid \delta) + \mathbb{E}[q_i^t \mid M](1 - ae^*(\delta, 0 \mid \delta))\right]
\end{aligned}$$

so that the first term of (8) is non-negative. Taken together, we have (8) $\geq 0$.

(ii). $e^*(\delta, 1 \mid \delta) > e^*(\delta, 0 \mid \delta)$.

From Lemma 5, it is shown that

$$e^*(\delta, 1 \mid \delta) > e^*(\delta, 0 \mid \delta) \Leftrightarrow [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I]) > [(1-\kappa)\mathbb{E}[q]](1-\mathbb{E}[q_i^t \mid M]).$$

This condition holds because $\kappa + (1-\kappa)(1-\mathbb{E}[q]) > (1-\kappa)\mathbb{E}[q]$ from Assumption 1 and $\mathbb{E}[q_i^t \mid I] \leq \mathbb{E}[q_i^t \mid M]$ from Lemma 4. Hence, $e^*(\delta, 1 \mid \delta) > e^*(\delta, 0 \mid \delta)$.

From (i) and (ii), we conclude that $V'(\delta) > 0$ as long as $V(\delta) \geq 0$. This directly implies that $\delta$ satisfying $V(\delta) = 0$ is unique if it exists.

Step 2. Now, we are ready to prove the proposition.

Case (i). There exists a unique $\delta$ satisfying $V(\delta) = 0$. Then, in the equilibrium, $\bar{\delta}$ is uniquely determined by the solution to $V(\delta) = 0$. Furthermore, $\bar{\delta} > 0$ because

$V(0) < 0$.

Case (ii). $V(\delta) < 0$ for any $\delta \in (0, \Delta)$. In this case, in the equilibrium, $\bar{\delta} = \Delta$.

From cases (i) and (ii), we obtain the desired result. $\square$

## A.5 Proof of Proposition 2 (ii)

Step 1. As a first step, we prove that $\frac{\partial V(\delta)}{\partial \sigma^2} < 0$ holds when $V(\delta) \geq 0$. Suppose that $V(\delta) \geq 0$ holds.

$$\begin{aligned}\frac{\partial V(\delta)}{\partial \sigma^2} =& b\left(\frac{\partial P^*}{\partial \sigma^2}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \sigma^2}(\delta, 0 \mid \delta)\right) \\ &+ \frac{\partial e^*(\delta, 1 \mid \delta)}{\partial \sigma^2}\frac{\partial U}{\partial e}(1, e^*(\delta, 1 \mid \delta)) - \frac{\partial e^*(\delta, 0 \mid \delta)}{\partial \sigma^2}\frac{\partial U}{\partial e}(1, e^*(\delta, 0 \mid \delta)) \\ =& b\left(\frac{\partial P^*}{\partial \sigma^2}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \sigma^2}(\delta, 0 \mid \delta)\right),\end{aligned}$$

where the second equality comes from the fact that

$$\frac{\partial U}{\partial e}(1, e^*(\delta, 1 \mid \delta)) = \frac{\partial U}{\partial e}(0, e^*(\delta, 0 \mid \delta)) = 0$$

because of the envelope theorem. Hence, it suffices to prove that

$$\frac{\partial P^*}{\partial \sigma^2}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \sigma^2}(\delta, 0 \mid \delta) < 0.$$

A simple calculation yields:

$$\begin{aligned}&\frac{\partial P^*}{\partial \sigma^2}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \sigma^2}(\delta, 0 \mid \delta) \\ &\quad= \psi\left\{[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\left[ae^*(\delta, 1 \mid \delta) + \mathbb{E}[q_i^t \mid I](1 - ae^*(\delta, 1 \mid \delta))\right]\right. \\ &\qquad \left. -(1-\kappa)\mathbb{E}[q]\left[ae^*(\delta, 0 \mid \delta) + \mathbb{E}[q_i^t \mid M](1 - ae^*(\delta, 0 \mid \delta))\right]\right\} \\ &\qquad \times \frac{\partial}{\partial \mathbb{E}[y_o^* \mid m = \omega]}\left(\frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}\right)\frac{\partial \mathbb{E}[y_o^* \mid m = \omega]}{\partial \sigma^2} \\ &\qquad + (1 - ae^*(\delta, 1 \mid \delta))\frac{\partial \mathbb{E}[q_i^t \mid I]}{\partial \sigma^2} - (1 - ae^*(\delta, 0 \mid \delta))\frac{\partial \mathbb{E}[q_i^t \mid M]}{\partial \sigma^2}. \qquad (9)\end{aligned}$$

Here, the last two terms are obviously negative because $\frac{\partial \mathbb{E}[q_i^t|I]}{\partial \sigma^2} < 0$ and $\frac{\partial \mathbb{E}[q_i^t|M]}{\partial \sigma^2} > 0$ holds. Furthermore, the first bracket is non-negative because $V(\delta) \geq 0$ is assumed. Hence, (9) is negative if and only if

$$\frac{\partial}{\partial \mathbb{E}[y_o^* \mid m = \omega]}\left(\frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}\right)\frac{\partial \mathbb{E}[y_o^* \mid m = \omega]}{\partial \sigma^2} \leq 0 \Leftrightarrow \frac{\partial \mathbb{E}[y_o^* \mid m = \omega]}{\partial \sigma^2} \geq 0$$

holds. Therefore, it suffices to show that $\frac{\partial \mathbb{E}[y_o^*|m=\omega]}{\partial \sigma^2} \geq 0$.

Prove this by contradiction. Suppose that $\frac{\partial \mathbb{E}[y_o^* | m=\omega]}{\partial \sigma^2} < 0$ holds for some $\sigma^2$ so that $\mathbb{E}[y_o^* \mid m = \omega]$ is decreasing in $\sigma^2$ for some $\sigma^2 \in [a, b]$ where $a, b > 0$. Then, from the derivation of $e^*(\delta, 1)$ in Lemma 5, this implies that $e^*(\delta, 1)$ is increasing in $\sigma^2 \in [a, b]$ for any $\delta$.[21] However, this further implies that $\mathbb{E}[y_o^* \mid m = \omega]$ is decreasing in $\sigma^2 \in [a, b]$, which is a contradiction.

Therefore, $\frac{\partial \mathbb{E}[y_o^* | m=\omega]}{\partial \sigma^2} \geq 0$, implying that (9) is negative.

Step 2. Now, we are ready to prove the proposition. Let $\bar{\delta}$ given $\sigma^2$ be $\bar{\delta}(\sigma^2)$ and $V(\delta)$ given $\sigma^2$ be $V(\delta \mid \sigma^2)$. Prove by contradiction. Suppose that $\bar{\delta}$ is non-increasing in $\sigma^2 \in [a, b]$ where $0 < a < b$. $V(\delta \mid \sigma^2) \geq 0$ holds for any $\delta (\geq \bar{\delta}(a))$ and $\sigma^2 \in [a, b]$. Hence, from Step 1, $0 = V(\bar{\delta}(a) \mid a) > V(\bar{\delta}(a) \mid b)$. This implies that $\bar{\delta}(b) > \bar{\delta}(a)$, which is a contradiction. Therefore, we obtain the desired result. □

## A.6 Proof of Proposition 3 (i)

We derive the value of $e^*(\delta, 1)$. For this, it is useful to obtain a formula of $e^*(\delta, 1)$ different from that in the proof of Lemma 5. Let $R_m := \mathbb{E}[p_i \mid m, y_1 = 1] - \mathbb{E}[p_i \mid m, y_1 = 0]$. Then, a straightforward calculation similar with that in the proof of Lemma 5 yields

$$e^*(\delta, 1) = \delta a b \psi R_1.$$

Hence, it suffices to derive the value of $R_1$.

Now,

$$\begin{aligned} R_1 =& \mathbb{E}[p_i \mid m, y_1 = 1] - \mathbb{E}[p_i \mid m, y_1 = 0] \\ =& \frac{[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = 1]} - \frac{[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\mathbb{E}[q_i^t \mid I]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = 1]}. \end{aligned} \tag{10}$$

Furthermore, since $\bar{\delta} = 0$,

$$\mathbb{E}[y_o^* \mid m = \omega] = a \int_0^{\Delta} e^*(\delta, 1) \frac{1}{\Delta} d\delta = a \int_0^{\Delta} \frac{\delta a b \psi R_1}{\Delta} d\delta = \frac{a^2 b \psi \Delta}{2} R_1;$$

[21] This holds because $\frac{\partial e^*(\delta,1)}{\partial \sigma^2}$ is given by

$$\delta a b \psi \bar{p} [\kappa + (1-\kappa)(1-\mathbb{E}[q])] \left[ \frac{\partial}{\partial \mathbb{E}[y_o^* \mid m = \omega]} \left( \frac{1}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]} \right) \frac{\partial \mathbb{E}[y_o^* \mid m = \omega]}{\partial \sigma^2} \right] \frac{\partial \mathbb{E}[q_i^t \mid I]}{\partial \sigma^2} < 0.$$

and $\bar{p} = \frac{1}{2-\varepsilon}$. By substituting them into (10), we obtain the following equation:

$$R_1 = \frac{[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}{1+(1-\varepsilon)\frac{a^2 b\psi\Delta}{2}R_1}$$
$$\Leftrightarrow (1-\varepsilon)\frac{a^2 b\psi\Delta}{2}R_1^2 + R_1 - [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I]) = 0.$$

That is, $R_1$ is given by $x \in [0,1]$ that is a solution to

$$(1-\varepsilon)\frac{a^2 b\psi\Delta}{2}x^2 + x - [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I]) = 0.$$

When $x = 0$, the left-hand side is negative so that

$$R_1 = \frac{-1+\sqrt{1+2(1-\varepsilon)a^2 b\psi\Delta[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}}{(1-\varepsilon)a^2 b\psi\Delta}.$$

By substituting this into $e^*(\delta,1) = \delta a b\psi R_1$, we obtain

$$e^*(\delta,1) = \frac{\delta}{(1-\varepsilon)a\Delta}\left(-1+\sqrt{1+2(1-\varepsilon)a^2 b\psi\Delta[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}\right).$$

Furthermore, this is obviously increasing in $\delta$. □

## A.7 Proof of Proposition 3 (ii)

The only difference from (i) is that the opportunistic type chooses $m \neq \omega$ if and only if $\delta < \bar{\delta}$. Hence, equation (10) itself does not change. Instead, $\mathbb{E}[y_o^* \mid m = \omega]$ and $\bar{p}$ change as follows:

$$\mathbb{E}[y_o^* \mid m = \omega] = a\int_{\bar{\delta}}^{\Delta} e^*(\delta,1)\frac{1}{\Delta-\bar{\delta}}d\delta = a\int_0^{\Delta}\frac{\delta a b\psi R_1}{\Delta-\bar{\delta}}d\delta = \frac{a^2 b\psi(\Delta+\bar{\delta})}{2}R_1; \tag{11}$$

$$\bar{p} = \frac{\Delta}{\Delta+(1-\varepsilon)(\Delta-\bar{\delta})}. \tag{12}$$

Based on them, we derive $e^*(\delta,1)$ and $e^*(\delta,0)$.

Step 1. We first derive $e^*(\delta,1)$. By substituting (11) and (12) into equation (10), we have

$$R_1 = \frac{\Delta[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}{\Delta+(1-\varepsilon)\frac{a^2 b\psi(\Delta^2-\bar{\delta}^2)}{2}R_1}$$
$$\Leftrightarrow (1-\varepsilon)\frac{a^2 b\psi(\Delta^2-\bar{\delta}^2)}{2}R_1^2 + \Delta R_1 - [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])\Delta = 0. \tag{13}$$

That is, $R_1$ is given by $x \in [0,1]$ that is a solution to

$$(1-\varepsilon)\frac{a^2 b\psi(\Delta^2-\bar{\delta}^2)}{2}x^2 + \Delta x - [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])\Delta = 0.$$

When $x = 0$, the left-hand side is negative so that

$$R_1 = \frac{-\Delta + \sqrt{\Delta^2 + 2(1-\varepsilon)a^2 b\psi\Delta(\Delta^2-\bar{\delta}^2)[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}}{(1-\varepsilon)a^2 b\psi(\Delta^2-\bar{\delta}^2)}.$$

By substituting this into $e^*(\delta,1) = \delta ab\psi R_1$, we obtain

$$\begin{aligned} e^*(\delta,1) =& \frac{\delta}{(1-\varepsilon)a(\Delta^2-\bar{\delta}^2)} \\ & \left(-\Delta + \sqrt{\Delta^2 + 2(1-\varepsilon)a^2 b\psi(\Delta^2-\bar{\delta}^2)[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}\right). \end{aligned}$$

Furthermore, this is obviously increasing in $\delta$.

Step 2. Next, we derive the value of $e^*(\delta,0)$. Given that $m = 0$, the opportunistic type's effort level in equilibrium is:

$$e^*(\delta,0) = \delta ab\psi R_0.$$

Hence, it suffices to derive the value of $R_0$. Now,

$$R_0 = \frac{(1-\kappa)\mathbb{E}[q](1-\mathbb{E}[q_i^t \mid M])\Delta}{\Delta + 0.5(1-\varepsilon)(\Delta^2-\bar{\delta}^2)a^2 b\psi R_1}. \tag{14}$$

Therefore, by using the value of $R_1$, we obtain

$$e^*(\delta,0) = \frac{2\delta ab\psi(1-\kappa)\mathbb{E}[q](1-\mathbb{E}[q_i^t \mid M])\Delta}{\Delta + \sqrt{\Delta^2 + 2(1-\varepsilon)a^2 b\psi(\Delta^2-\bar{\delta}^2)[\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i^t \mid I])}}.$$

Furthermore, this is obviously increasing in $\delta$.

Step 3. Lastly, we prove that $\lim_{\delta \nearrow \bar{\delta}} e^{**}(\delta) < \lim_{\delta \searrow \bar{\delta}} e^{**}(\delta)$. This holds because $e^*(\delta,1) > e^*(\delta,0)$ from the proof of Proposition 2. □

## A.8 Proof of Proposition 4

(i) is straightforward because $e^{**}$ is decreasing in $\mathbb{E}[q^t \mid I]$ and higher $\sigma^2$ reduces $\mathbb{E}[q^t \mid I]$ from Lemma 4. Thus, we focus on (ii).

Step 1. First, from Proposition 3, $\lim_{\delta \nearrow \bar{\delta}} e^{**}(\delta) < \lim_{\delta \searrow \bar{\delta}} e^{**}(\delta)$. Furthermore, $\bar{\delta}(a+\varepsilon) > \bar{\delta}(a)$ from Proposition 2. Combining them yields the fact that $e^{**}(\delta \mid a+\varepsilon) < e^{**}(\delta \mid a)$ for $\delta \in [\bar{\delta}(a), \bar{\delta}(a+\varepsilon))$.

Step 2. Next, we prove that $e^{**}(\delta \mid a+\varepsilon) > e^{**}(\delta \mid a)$ for $\delta \in [\bar{\delta}(a+\varepsilon), \Delta)$. For this, it suffices to prove that $e^{**}(\delta)$ is increasing in $\sigma^2$ for $\delta > \bar{\delta}(a+\varepsilon)$. In particular, we will prove that

$$\frac{dR_1(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} = \frac{\partial R_1}{\partial \bar{\delta}}(\sigma^2, \bar{\delta}(\sigma^2))\frac{\partial \bar{\delta}}{\partial \sigma^2} + \frac{\partial R_1}{\partial \sigma^2}(\sigma^2, \bar{\delta}(\sigma^2)) > 0. \tag{15}$$

If this holds, $e^*(\delta, 1)$ is increasing in $\sigma^2$ because $e^*(\delta, 1) = \delta ab\psi R_1$.

As a first step, we derive $\frac{\partial R_1}{\partial \bar{\delta}}$. By applying the implicit function theorem with respect to $\bar{\delta}$ for equation (13), we obtain the following:

$$\frac{\partial R_1}{\partial \bar{\delta}} = \frac{(1-\varepsilon)a^2 b\psi R_1^2 \bar{\delta}}{(1-\varepsilon)(\Delta^2 - \bar{\delta}^2)a^2 b\psi R_1 + \Delta} > 0.$$

Furthermore, $\frac{\partial \bar{\delta}}{\partial \sigma^2} > 0$ holds from Proposition 2.

Next, we derive $\frac{\partial R_1}{\partial \sigma^2}(\sigma^2, \bar{\delta}(\sigma^2))$. By applying the implicit function theorem with respect to $\sigma^2$ for equation (13), we obtain the following:

$$\frac{\partial R_1}{\partial \sigma^2} = -\frac{\Delta[\kappa + (1-\kappa)(1-\mathbb{E}[q])] \times \frac{\partial \mathbb{E}[q_i^t|I]}{\partial \sigma^2}}{(1-\varepsilon)(\Delta^2 - \bar{\delta}^2)a^2 b\psi R_1 + \Delta} = \frac{\Delta\lambda(1-\kappa)}{(1-\varepsilon)(\Delta^2 - \bar{\delta}^2)a^2 b\psi R_1 + \Delta} > 0.$$

Combining them yields (15), which completes the proof of Step 2.

Step 3. The last step is to show that $e^{**}(\delta \mid a+\varepsilon) < e^{**}(\delta \mid a)$ for $\delta \in (0, \bar{\delta}(a))$. For this, it suffices to prove that $e^{**}(\delta)$ is decreasing in $\sigma^2$ for $\delta < \bar{\delta}(a)$. In particular, we will prove that

$$\frac{dR_0(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} = \frac{\partial R_0}{\partial R_1}\frac{dR_1(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} + \frac{\partial R_0}{\partial \sigma^2} < 0. \tag{16}$$

If this holds, $e^*(\delta, 0)$ is decreasing in $\sigma^2$ because $e^*(\delta, 0) = \delta ab\psi R_0$.

First, from equation (14), it is straightforward that $\frac{\partial R_0}{\partial R_1} < 0$. Furthermore, from Step 2, $\frac{dR_1(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} > 0$. Hence,

$$\frac{\partial R_0}{\partial R_1}\frac{dR_1(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} < 0.$$

Second, $R_0$ is decreasing in $\mathbb{E}[q_i^t \mid M]$ and $\mathbb{E}[q_i^t \mid M]$ is increasing in $\sigma^2$. Hence, $\frac{\partial R_0}{\partial \sigma^2} < 0$.

Combining them yields (16), which completes the proof of Step 3.

From Steps 1-3, we obtain the desired result. $\square$

## A.9 Proof of Proposition 5

Let $p_i^I$ be voter $i$'s belief on the incumbent's type, whereas let $p_i^C$ be his belief on the challenger's type. It can be shown that $p_i^I$ depends on only $(m_I, \tilde{y}_i)$, whereas $p_i^C$ depends on only $m_C$.

By using the procedure same as in the main analysis, we obtain the expected probability of the incumbent obtaining a majority of votes given $(m_I, m_C, e_1)$ as:

$$\frac{1}{2} + \psi \left( \mathbb{E}[p_i^I \mid m_I, e_1] - \mathbb{E}[p_i^C \mid m_C] \right). \tag{17}$$

(i). The incumbent's objective is to maximize (17). Since the incumbent's choice variable influences only $\mathbb{E}[p_i^I \mid m_I, e_1]$, it is equivalent to maximizing $\mathbb{E}[p_i^I \mid m_I, e_1]$. Hence, the incumbent's decision problem is reduced to that same as in the main analysis. Therefore, the incumbent's equilibrium strategy remains the same.

(ii). The challenger's objective is to minimize (17), which is equal to maximizing $\mathbb{E}[p_i^C \mid m_C]$. As in the main analysis, when $m = 1$, informed voters think that the incumbent supports the truth so that her reputation becomes $\bar{p}$, whereas misinformed voters think that the incumbent denies the truth so that her reputation becomes zero. Hence,

$$\mathbb{E}[p_i^C \mid m_C = 1] = [\kappa + (1-\kappa)(1 - \mathbb{E}[q])]\bar{p}.$$

Note that $\mathbb{E}[m^*(\delta, 1)]$ in the definition of $\bar{p}$ is replaced by $\mathbb{E}[m_C^*(\delta, 1)]$. Similarly, when $m = 0$, the incumbent's reputation becomes zero among informed voters, but becomes $\bar{p}$; thus,

$$\mathbb{E}[p_i^C \mid m_C = 0] = (1-\kappa)\mathbb{E}[q]\bar{p}.$$

From Assumption 1, $\mathbb{E}[p_i^C \mid m_C = 1] > \mathbb{E}[p_i^C \mid m_C = 0]$, implying that the challenger never denies the truth strategically. □