Misinformation in Representative Democracy: The Role of Heterogeneous Confirmation Bias*

Satoshi Kasamatsu[†] Daiki Kishishita[‡]

First version: November, 2022 Latest version: August 3, 2024

Abstract

We investigate why politicians promote false claims supported by only a minority of voters and how it impacts representative democracy. We construct an electoral accountability model where an incumbent politician decides whether to deny the truth about an issue and exerts effort to produce public goods. The key is heterogeneity of confirmation bias that leads people to persistently accept misinformation. Under heterogeneous confirmation bias, promoting misinformation enables the incumbent to weaken electoral punishment for low performance. Therefore, low-competent politicians may deny the truth. Moreover, social media platforms amplify the heterogeneity of confirmation bias, thereby inducing politicians' misinformation.

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

JEL classification: D72, D83, D91

*We would like to thank Navin Kartik (an editor), anonymous referees, Yasushi Asako, Yuki Atsusaka, Yu Awaya, Emre Aytimur, Kentaro Fukumoto, Chiaki Hara, Kazumi Hori, Yukari Iwanami, Atsushi Kajii, Hirofumi Miwa, Takeshi Miyazaki, Daisuke Nakajima, Masayuki Odora, Hitoshi Sadakane, Susumu Sato, Tadashi Sekiguchi, Yasuhiro Shirata, Naoko Taniguchi, Takuro Yamashita, Michael Zierhut, the participants of the 2022 Summer Workshop on Economic Theory in Otaru, the 2022 Annual Meeting of Japan Public Choice Society at Nagoya City University, the 2023 Annual Meeting of Japanese Association of Electoral Studies at University of Tokyo, the 2023 Spring Meeting of Japanese Economic Association (online), and the 2023 Summer Meeting of the Japanese Society for Quantitative Political Science at Fukuoka University, and the seminar participants at Kyoto University, Doshisha University, Ryukoku University, and Hitotsubashi University 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. This manuscript was previously circulated under the title "Strategic Misinformation: The Role of Heterogeneous Confirmation Bias."

[†]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 spread of misinformation such as fake news and conspiracy theories is a growing concern in democracies (Jerit and Zhao, 2020; Nyhan, 2020).[1] A notable feature of this spread is that even politicians holding office support false claims. Examples include former president Donald Trump in the US and former president Jair Bolsonaro in Brazil.[2] To address the prevalence of misinformation among elected politicians,[3] institutions such as PolitiFact in the US conduct fact-checking of their claims. While the literature has paid much attention to the spread of misinformation among citizens, less is known about why politicians promote misinformation and how it affects representative democracy, especially the accountability mechanism of elections. The present study addresses this issue from the perspective of politicians' electoral incentives.

As politicians' electoral incentives depend on the demand side, a starting point is to understand why voters persistently believe in misinformation. The literature on political psychology has analyzed this issue and identified psychological biases as a key determinant of persistently accepting misinformation. In other words, to reach a desired conclusion, people process 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). Hence, even after facing factual information, they may not correct their false hypotheses.[4] For instance, Nyhan and Reifler (2010) find that correcting misinformation that Saddam Hussein had weapons of mass destruction did not mitigate, but rather reinforced this misinformation among conservatives. Specifically, 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).[5]

[1]Note that not all conspiracy theories are demonstrably false (see Pantazi, Hale and Klein (2021) for the conceptions of these terms).

[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). For information about President Bolsonaro, see Ricard and Medeiros (2020).

[3]Lasser et al. (2022) examine whether members of the parliament in the US, UK, and Germany share untrustworthy information on Twitter. A certain share of elected politicians does share untrustworthy information, while it is also the case that many politicians share only trustworthy information.

[4]To be fair, the observed empirical patterns may be explained by Bayesian learning (see Little (2022)).

[5]Faia et al. (forthcoming) find confirmation bias as a cause of biased updating about COVID-19.

However, this demand-side factor cannot readily explain the political supply side. Even in the presence of psychological biases, the majority of voters would know the truth. For example, although COVID-19 conspiracy theories are widespread, only 31% of US citizens believe in the conspiracy theory that the virus was purposefully created and spread as of 2020 (Uscinski et al., 2020). As long as only a minority of voters believe in misinformation, politicians should support the truth to win an election.[6]

We argue that the key to resolving this paradox is the heterogeneity in the degree of confirmation bias. Recently, several empirical studies have examined individual differences in motivated reasoning, such as confirmation bias. Arceneaux and Vander Wielen (2017) find that a difference in the need for cognition and the need for affect determines the level of political motivated reasoning and show that these factors are heterogeneous across citizens.[7] Hohnsbehn, Urschler and Schneider (2022) also find that people with higher trait ambivalence show less confirmation bias. In this study, we argue that the extent of such individual differences in confirmation bias creates room for politicians' use of misinformation.

For this purpose, we extend the canonical two-period political agency model à la Besley (2006). Our model comprises three types of players: the incumbent politician, a challenger, and a continuum of voters. In period 1, the incumbent exerts an effort such as monitoring bureaucrats, which influences her performance regarding public goods provision. After observing the performance, voters decide whether to reelect the incumbent or elect the challenger based on probabilistic voting. In period 2, the elected politician exerts an effort. Politicians are either the ethical type (i.e., sincerely exerting the highest effort) or the opportunistic type (i.e., caring for effort cost and reelection incentive). The effort cost for the opportunistic type is uniformly distributed, which is interpreted as the degree of (in)competence of this type of politician. Each politician's type and her effort cost are unknown to the voters.

We introduce the communication stage into this standard model. At the beginning of the game, players face a controversial issue (e.g., the safety of a COVID-19 vaccine) represented by a binary state of the world ($\omega \in \{0, 1\}$). We assume that state 1 is true. Some voters have a prior belief that the state is likely to be one, whereas others have a prior that the state is likely

[6]There are two potential explanations for this paradox. First, if the acceptance of misinformation is positively associated with the turnout rate, misinformed voters may constitute a majority among voters who actually vote. Ardèvol-Abreu, Gil de Zuniga and Gámez (2020) empirically examine whether the acceptance of conspiracy theories positively correlates with formal political participation such as voting and find the opposite. Second, politicians may deny the truth because they want to do so ideologically without caring about an election. However, Nyhan and Reifler (2015) empirically confirm that fact-checking reduces politicians' misinformation because they care about reputation risk. Either of these two explanations is rejected.

[7]In addition, Taber and Lodge (2006) and Kahan et al. (2017) find that a difference in political knowledge and science curiosity correlates with the individual level of motivated reasoning. Likewise, Enke and Zimmermann (2019) show the individual heterogeneity of correlation neglect, another well-known type of biased updating.

to be zero. The factual information from experts that state 1 is true is reported as a perfect signal of the true state. However, some voters who initially believed that state 0 is more likely to be true misinterpret this information and reinforce their initial argument because of confirmation bias. Consequently, some voters remain *misinformed* (i.e., wrongly believe that state 0 is true). On the other hand, the others correctly understand that the truth is state 1 (i.e., they are *informed*). We assume that the majority of voters are informed. Building on Rabin and Schrag (1999), we model voters' confirmation bias such that, with probability q_i, voter i misinterprets a belief-inconsnant signal as a signal consistent with their prior belief. q_i can be regarded as the degree of voter i's confirmation bias. In this environment, the incumbent politician chooses whether to support the truth (i.e., state 1) or deny it, which is a cheap-talk message. The sincere type is assumed to convey the truth. After this stage, the two-period election model described above is played.

We show that an opportunistic-type incumbent may have the incentive to deny the truth supported by the majority of voters when her competence level in producing public goods is low. Particularly, a low-competent politician has this incentive if and only if the degree of confirmation bias is sufficiently heterogeneous across voters. Notably, the incumbent neither misunderstands the truth nor wants voters to believe in the misinformation. In our model, politicians know the truth, and their payoff is defined as the benefit of reelection minus the effort cost. Nonetheless, to win an election, a low-competent incumbent denies the truth supported by the majority of voters under sufficiently heterogeneous confirmation bias.

The mechanism is as follows. First, suppose the degree of confirmation bias is common among voters. On the one hand, when the incumbent supports the truth, misinformed voters think that the incumbent argues the false claim; thus, the probability of the incumbent being ethical (i.e., the incumbent's reputation) becomes zero among them. As a result, the incumbent's political support base consists of informed voters when supporting the truth. On the other hand, when the incumbent denies the truth, her reputation becomes zero among informed voters, but it is improved among misinformed voters because misinformed voters incorrectly think that the incumbent claims the truth. Hence, when denying the truth, her political support base consists of misinformed voters. A larger political support base increases the chance of reelection. Therefore, the incumbent never strategically denies the truth supported by the majority of voters.

However, a part of this logic fails when confirmation bias is heterogeneous across voters. The key is that the distribution of confirmation bias diverges between informed and misinformed voters. 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. With higher confirmation bias, voters are more likely to misinterpret the incumbent's low performance in producing public goods as high once they thought that the incumbent would be ethical. Thus, the higher the bias the less likely voters are to replace the low-performed incumbent. These properties imply that targeting misinformed voters as a political support base enables the incumbent to weaken electoral punishment.

Therefore, the incumbent faces the pros and cons of denying the truth in the presence of heterogeneous confirmation bias. The cost is that her political support base shrinks, and the benefit is that her political support base has a higher confirmation bias; hence, electoral punishment becomes weaker. The performance of low-competent politicians tends to be low. Therefore, the benefit of denying the truth is large enough for them to overturn the cost. As such, low-competent politicians deny the truth under heterogeneous confirmation bias. Examining three cases of populists in government, Pirro and Taggart (2023) argue that incumbent populists tend to spread conspiracy theories. Our results align with this argument because as political outsiders, populists may lack high competence in daily policymaking.

Importantly, this analysis shows that politicians' use of misinformation has a hidden negative effect on representative democracy. Scholars would believe that politicians should respect the truth because denying it induces ordinary citizens to hold wrong beliefs, which may hurt society. While we share the same concern, our finding uncovers another hidden cost of misinformation. In our model, low-competent politicians deny the truth to weaken the electoral punishment for low performance. As such, strategic misinformation enables low-competent politicians to weaken the fundamental role of an election as a device for disciplining politicians, which reduces their effort level.

This negative effect on electoral accountability differs from conventional wisdom. Typically, the negative impact we associate with politicians' use of misinformation is that when candidates spread new false information, voters adopt incorrect beliefs, thus undermining accountability. However, in practice, misinformation is often already widespread before politicians endorse it. Politicians tend to reinforce existing misinformation rather than create it. Additionally, politicians may spread misinformation on topics unrelated to elections. Our findings indicate that the incumbent's use of misinformation is harmful even if the misinformation is already widespread and not directly related to the election.

Overall, our results yield two testable predictions in addition to resolving the paradox of politicians' use of misinformation. First, less competent politicians are more likely to spread misinformation. Second, spreading misinformation reduces the incumbent's effort in providing public goods. While several empirical studies have analyzed the effect of politicians' partisanship on their propensity to spread misinformation (e.g., Lasser et al., 2022) and the impact of misinformation on electoral outcomes (e.g., Barrera et al., 2020; Malzahn and Hall, 2023), our two predictions have not yet been empirically examined. Our model offers these new predictions, which can be tested in future research. In addition, our model provides several further implications for empirical research, which will be discussed in Section 6.

Finally, we discuss an application of the proposed model. In practice, confirmation bias is a joint product of the psychological factor and the information environment. Specifically, social media platforms can facilitate the spread of fake news and create echo-chambers (e.g., Del Vicario et al., 2016; Allcott and Gentzkow, 2017; Levy, 2021), which amplify the confirmation bias resulting from the psychological factor. By extending our model, we show that social media platforms lead to a greater variance in confirmation bias among individuals. As a result of this effect, the presence of social media platforms encourages politicians' strategic use of misinformation. This provides a new rationale for the importance of efforts to combat fake news and echo chambers on social media. Such attempts are indispensable not only because they prevent the spread of misinformation among citizens but also because they reduce the incumbent's strategic use of misinformation weakening electoral accountability.

Related literature: First, this study contributes to the growing literature that examines the effect of voters' confirmation bias on elections (Levy and Razin, 2015; Lockwood, 2017; Millner, Ollivier and Simon, 2020; Little, Schnakenberg and Turner, 2022).[8] Lockwood (2017) finds its negative effect on electoral accountability using Rabin and Schrag (1999)'s framework, similar to our study;[9] and Little, Schnakenberg and Turner (2022) develop a general framework for motivated reasoning and apply it to the analysis of electoral accountability. While our study is motivated by these studies, neither of them addresses the issue of misinformation nor heterogeneous confirmation bias.[10]

Second, the literature on propaganda explores ways wherein politicians can induce a particular action of voters by manipulating information (e.g., Glaeser, 2005; Gehlbach and

[8]Subsequent works on modeling confirmation bias include 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.

[9]He analyzes two models: the first is where the optimal policy and performance are observable such as the incumbent's effort choice, and the second is where these are unobservable. The former one is adopted by our study. For the latter model, he shows that confirmation bias reduces the incumbent's pandering incentive.

[10]Kartal and Tyran (2022) analyze information aggregation in a voting game with overconfidence and misinformation. Politicians' incentives are not examined because it concerns the efficiency of direct democracy.

Sonin, 2014; Little, 2017; Guriev and Treisman, 2020; Horz, 2021; Bräuninger and Marinov, 2022). In these studies, politicians manipulate information directly related to their office-holding or policymaking (e.g., an authoritarian regime manipulates information about the regime's quality to prevent protests). Conversely, our model assumes that voters' belief about what is the truth is not relevant for politicians. Nonetheless, politicians deny the truth because it indirectly influences reelection through its effect on electoral accountability. We complement the existing literature by presenting another incentive for misinformation.

Third, this study relates to research showing that unpopular policies may increase the likelihood of electoral victory. Previous studies have identified several mechanisms, including cultivating a base of loyal supporters (Glaeser, Ponzetto and Shapiro, 2005), signaling a commitment to high-quality policymaking (Carrillo and Castanheira, 2008), demonstrating a politicians' character (Kartik and McAfee, 2007), demonstrating that a politician's preference is aligned with that of voters (Acemoglu, Egorov and Sonin, 2013), and investing in issues that may become popular in the future (Eguia and Giovannoni, 2019). Our study contributes to the literature by introducing a new mechanism driven by confirmation bias. The work of Grillo and Prato (2023) is particularly related because they demonstrate that even when most citizens value democracy, democratic backsliding occur under voters' reference-dependent preferences. Similar to our study, their work considers psychological bias, but do not find the role of heterogeneous psychological bias. Additionally, the work of Kartik and Van Weelden (2019*a*) is relevant as it shows that a politician's winning probability may be enhanced by conveying the message that she is of a bad type. In their model, reputation concerns distort post-election policymaking through pandering. As a result, voters may prefer a candidate known to be bad over one whose type is uncertain. Unlike their model, in our model, the more likely a politician is to be bad, the more negative the impact on voters. Nonetheless, we find that unpopular messages can increase the likelihood of electoral victory.

In contrast to these studies, our model generates novel testable predictions: (i) low-competence politicians are more likely to claim misinformation, and (ii) claiming misinformation decreases the incumbent's effort in providing public goods. Existing studies do not yield these predictions because their models lack the consideration of politicians' effort choices regarding public goods provision in addition to their decisions on supporting unpopular policies and opinions.[11] These predictions are crucial as they illuminate how the strategic use of misinformation by politicians in one area can undermine policymaking in others.

[11] Kartik and Van Weelden (2019*a*) consider policymaking after politicians convey messages, but they do not obtain these predictions.

Fourth, as an application, we analyze the impact of social media. Therefore, our study is related to the literature on social media and electoral accountability. Li and Hu (2023) and Kocak and Kıbrıs (2023) analyze the impact of social media on media capture. Li, Raiha and Shotts (2022) investigate whether alternative media serving as watchdogs of mainstream media improve electoral accountability. Furthermore, Li and Hu (2023) demonstrate that personalized news bias voters' opinions but may enhance electoral accountability. In contrast, we focus on a new aspect of social media –its role in increasing the variance in confirmation bias – and analyze its impact on a new outcome – politicians' strategic use of misinformation.

Finally, the literature has paid limited attention to politicians' use of misinformation. To our knowledge, only two exceptions exist. First, Szeidl and Szucs (2022) investigate politicians' use of a conspiracy theory. Their key assumption is that voters believing in the conspiracy come to distrust the intellectual elite and not to believe the media's news. Thus, to weaken electoral punishment based on news reporting, the incumbent with low quality persuades voters of the conspiracy theory. Second, Grossman and Helpman (2023) examine the strategic use of misinformation in electoral competition, where politicians can send misinformation about the policy environment and parties' positions. They show that policy divergence can result.

Our study makes three main contributions compared with these two studies. First, we demonstrate that misinformation does not have to be conspiracy theories connected with distrusting elites or electoral issues: simply being false is enough for it to be utilized for political purposes. Second, our model allows the impact of denying the truth to be different across the electorate: it increases the popularity of misinformed voters while decreasing the popularity of informed voters. While this is reasonable, the existing two studies do not account for these differing responses: Szeidl and Szucs (2022)'s model contains only a representative voter, and Grossman and Helpman (2023) assume that conveying misinformation does not undermine the popularity among informed voters because microtargeting is possible in campaigns. In contrast, our model shows that politicians may still employ misinformation even if it reduces their popularity among the majority of voters. Finally, the literature on political psychology has devoted considerable attention to the role of confirmation bias in the acceptance of misinformation. While the existing two studies do not introduce this bias, our model demonstrates that the heterogeneity of this bias is crucial in the strategic use of misinformation.

2 Motivating Example

We start with presenting a simplified example, which highlights how the heterogeneity of confirmation bias gives the incumbent politician an incentive to deny the truth.

Suppose that an incumbent politician faces an election with a continuum of voters. We will use the pronoun "she" for candidates and "he" for voters. Voters face a controversial issue whose truth is uncertain (e.g., whether taking a COVID-19 vaccine is safe): $\omega \in \{0, 1\}$. Although ω is uncertain for voters, the incumbent knows its value. Voters have heterogeneous prior beliefs on the value of ω. We suppose that a fraction κ of voters have a prior of $\Pr(\omega = k) = \bar{\alpha} \in (0.5, 1)$, whereas a fraction $1 - \kappa$ of voters have a prior of $\Pr(\omega = k) = \underline{\alpha} \in (0, 0.5)$.[12] Without loss of generality, we focus on the case wherein $\omega = 1$. 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 COVID-19 vaccination, s is a statement by the World Health Organization. To analyze misinformation, we must consider a situation wherein the truth is evident under rational reasoning. Otherwise, the truth is objectively uncertain. Thus, s is assumed to be perfect (i.e., $s = \omega$ with probability one). As $\omega = 1$ is assumed, $s = 1$.

After the signal s, the incumbent sends a cheap-talk message about ω: $m \in \{0, 1\}$. $m = \omega$ (resp. $m \neq \omega$) represents confirming the truth (resp. denying the truth). As $\omega = 1$ is assumed, $m = 1$ (resp. $m = 0$) represents confirming the truth (resp. denying the truth). In addition to communicating about the controversial issue with voters, the incumbent is involved in daily policymaking. Just before the election, voters observe the incumbent's performance on policymaking: $y \in \{0, 1\}$.

The incumbent is either the ethical type or the opportunistic type with an equal probability. The ethical type tells the truth, and her performance on policymaking is always high (i.e., $y = 1$). On the other hand, the opportunistic type's performance in policymaking is always low (i.e., $y = 0$), and she strategically decides whether to tell the truth to maximize the election probability. Let p_i be voter i's subjective probability of the incumbent being ethical at the time of the election. The incumbent's reelection probability is equal to $R := \int p_i di$, the average reputation of the incumbent.

Lastly, we introduce confirmation bias into this model. When facing a belief-consonant signal, he literally interprets the signal, but when facing a belief-inconsonant signal, he misinterprets the signal as the opposite signal with probability $q_i \in [0, 1]$. Then, q_i captures the extent of confirmation bias of voter i.

[12] Each voter does not know the distribution of prior beliefs because otherwise, they can infer the value of ω from the distribution.

We examine the existence of an equilibrium where the opportunistic type denies the truth.

Homogeneous confirmation bias: Suppose that every voter's extent of confirmation bias is the same: $q_i = \frac{1}{2}$ for every voter. For κ fraction of voters, the public signal $s = 1$ is a belief-consonant signal; thus, they literally interpret this signal and learn that $\omega = 1$. On the other hand, for $1 - \kappa$ fraction of voters, $s = 1$ is a belief-inconsonant signal; thus, they misunderstand that $s = 0$ and learn that $\omega = 0$ with probability $q_i = \frac{1}{2}$. Therefore, $\kappa + (1-\kappa)\frac{1}{2}$ fraction of voters is informed, whereas $(1-\kappa)\frac{1}{2}$ of voters is misinformed.

First, consider the case where the incumbent tells the truth (i.e., $m = 1$). Informed voters correctly think that the incumbent tells the truth so that the subjective probability of the incumbent being ethical becomes one. On the other hand, misinformed voters believe that $\omega = 0$; thus, they interpret $m = 1$ as indicating that the incumbent is opportunistic. As a result, the subjective probability of the incumbent being ethical becomes zero among them. After this communication, the performance in policymaking is realized. The performance of the opportunistic type is always low, which should make the incumbent's reputation zero. However, voters may misinterpret the low performance as the high performance as a result of confirmation bias. Specifically, informed voters misinterpret the low performance as the high performance with probability $q_i = \frac{1}{2}$ because they believe that the incumbent is ethical. Consequently,

$$R = \underbrace{\left[\kappa + (1-\kappa)\frac{1}{2}\right]}_{\text{size of support base}} \cdot \underbrace{\frac{1}{2}}_{\text{average confirmation bias}}. \quad (1)$$

Second, consider the case where the incumbent denies truth (i.e., $m = 0$). In this case, the incumbent's support base consists of misinformed voters. Furthermore, misinformed voters misinterpret that the performance is high with probability $q_i = \frac{1}{2}$. Therefore,

$$R = \left[(1-\kappa)\frac{1}{2}\right] \cdot \frac{1}{2}. \quad (2)$$

As long as the majority of voters are informed (i.e., $\kappa + (1-\kappa)\frac{1}{2} > \frac{1}{2}$), (1)>(2) holds; that is, claiming the truth is better in terms of the reelection probability. Therefore, the opportunistic type tells the truth so that there is no equilibrium that the opportunistic type denies the truth.

In this setting, when denying the truth, every informed voter stops supporting the incumbent. In practice, however, a certain fraction of informed voters may continue to support the incumbent under strong partisanship. However, even in such a case, the same result would hold as long as the distribution of partisanship is independent of whether voters are informed or not.

Heterogeneous confirmation bias: This result can be reversed in the presence of heterogeneous confirmation bias. Suppose that $q_i = 0$ with probability a half and $q_i = 1$ with probability a half. For κ fraction of voters, $s = 1$ is a belief-consonant signal; thus, they literally interpret this signal and learn that $\omega = 1$. On the other hand, among $1 - \kappa$ fraction of voters, those with $q_i = 0$ correctly learn that $\omega = 1$, but those with $q_i = 1$ wrongly learn that $\omega = 0$. Therefore, $\kappa + (1-\kappa)\frac{1}{2}$ fraction of voters is informed, whereas $(1-\kappa)\frac{1}{2}$ of voters is misinformed. A striking property is that the average confirmation bias differs across informed and misinformed voters. It is $\frac{\kappa\frac{1}{2}}{\kappa+(1-\kappa)\frac{1}{2}} = \frac{\kappa}{1+\kappa}$ among informed voters, whereas it is one among misinformed voters. That is, the average confirmation bias is higher among misinformed voters than among informed voters.

Consider the case where $m = 1$. As before, the support base consists of informed voters. Furthermore, fraction $\frac{\kappa}{1+\kappa}$ of informed voters misinterpret the low performance as the high performance. Consequently,

$$R = \left[\kappa + (1-\kappa)\frac{1}{2}\right] \cdot \frac{\kappa}{1+\kappa} = \frac{\kappa}{2}. \tag{3}$$

Second, consider the case where $m = 0$. The support base consists of misinformed voters. Furthermore, misinformed voters always misinterpret the low performance as the high performance because their q_i is equal to one. Therefore,

$$R = (1-\kappa)\frac{1}{2}. \tag{4}$$

When (3) $<$ (4) $\Leftrightarrow \kappa < \frac{1}{2}$, there is an equilibrium that the opportunistic type denies the truth even if the majority of voters are informed . Key is that the reelection probability is determined by the size of the support base average confirmation bias among the support base. In terms of the size, attracting informed voters is better. However, in the presence of heterogeneous confirmation bias, attracting misinformed voters is better in terms of average confirmation bias. Therefore, politicians may deny the truth.

3 Model

Our full model is an extension of the standard two-period electoral accountability model à la Besley (2006). It comprises three types of players: the incumbent politician, the challenger,

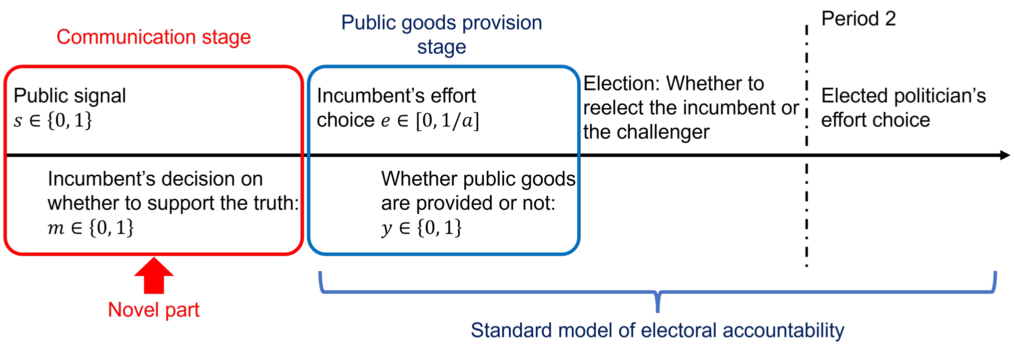

Figure 1: Timing of the Game

and the continuum of voters. At the beginning of period 1, the incumbent politician communicates her opinion about a controversial issue. This communication stage is exactly the same as in Section 2. Hereafter, without loss of generality, we focus on the case wherein $\omega = 1$. After the communication stage, the incumbent chooses her effort level for the provision of public goods. Then, voters vote for either the incumbent or the challenger. The candidate obtaining the majority of votes holds office and exerts an effort in period 2. The details are illustrated in Figure 1.

3.1 Stage of Public Goods Provision

In period $t \in \{1, 2\}$, the politician holding office exerts an effort to produce public goods: $e_t \in [0, 1/a]$, where $a > 0$. The effort can be interpreted as monitoring bureaucrats. In period 1, this is after the communication stage. Effort influences whether public goods are successfully provided (i.e., elected politician's performance). In particular, the elected politician's performance is given by $y_t \in \{0, 1\}$, where $y_t = 1$ represents the successful provision of public goods (i.e, high performance). The probability of $y_t = 1$ is given by $z(e_t) := ae_t$. Therefore, a higher effort increases the probability of public goods provision, and performance is always high when the exerted effort is maximum (i.e., $e_t = 1/a$). The public goods can be interpreted as high economic growth when we consider elections of the president and governors, while it can be interpreted as pork-barrel to the elected district when we consider elections of legislators.

3.2 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 as much effort for public goods provision as possible (i.e., $e = 1/a$). In Section 6, we analyze the case where the ethical type strategically behaves to maximize voters' welfare.

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 when the politician is holding office in period t. Here, $b > 0$ represents the office-holding benefit, whereas $e_t^2/(2\delta)$ represents the effort cost. δ is the politician's competence because higher δ reduces the effort cost. We allow δ to vary between politicians. Specifically, δ follows a uniform distribution $U(0, \Delta)$, where $\Delta > 0$.[13] To exclude implausible off-path beliefs, we also assume that with probability ε, this type of politician non-strategically denies the truth and exerts no effort for public goods provision (i.e., $m \neq \omega$ and $e = 0$). ε can be an arbitrarily small but positive value.

Voters do not know each politician's type, whereas each politician knows her own type. The value of δ is also known to the politician but unknown to voters. The prior probability of a politician being ethical is assumed to be 0.5.

3.3 Election

After observing the incumbent's performance, voters face an election at the beginning of period 2. Because our interest is the impact of misinformation on electoral accountability, we consider the setup where voters share identical preferences for the public goods provision (e.g., Besley, 2006): that is, every voter prefers the higher level of public goods provision. Specifically, voters' payoff from public goods provision in period t is given by y_t. Hence, the expected payoff in period 2 for voter i when reelecting the incumbent (resp. electing a challenger) is $\mathbb{E}_i[y_2 \mid m, \tilde{y}_{1i}, \text{Incumbent}]$ (resp. $\mathbb{E}_i[y_2 \mid m, \tilde{y}_{1i}, \text{Challenger}]$), where $\tilde{y}_{1i}$ is the value of y_1 perceived by voter i. We suppose probabilistic voting (Lindbeck and Weibull, 1987; Dixit and Londregan, 1996): voter i votes for the incumbent if and only if

$$\mathbb{E}_i[y_2 \mid m, \tilde{y}_{1i}, \text{Incumbent}] + \zeta_i + \eta \geq \mathbb{E}_i[y_2 \mid m, \tilde{y}_{1i}, \text{Challenger}]. \tag{5}$$

[13]This heterogeneity on competence is not crucial in that the incumbent may deny the truth even if we assume that δ is deterministic and takes a small value. We introduce it because it allows us to examine how the incumbent's message and effort level differs across different competence levels (Propositions 2, 3, and 11).

In addition to the utility derived from public goods provision, various stochastic factors influence voting behavior. These factors are formulated using ζ_i and η. ζ_i is an idiosyncratic shock specific to voter i, which follows a uniform distribution $U[-1/(2\gamma), 1/(2\gamma)]$, whereas η is an aggregate shock that is common across voters, which follows a uniform distribution $U[-1/(2\psi), 1/(2\psi)]$. $\gamma \in \left(0, \frac{\psi}{2\psi+1}\right]$ and $\psi \in (0, 0.5)$ are assumed to guarantee that their votes are always probabilistic.

To see more, let p_i be voter i's subjective probability of the incumbent being ethical at the time of election. In period 2, the opportunistic type chooses the lowest level of effort ($e_2 = 0$) because there is no reelection concern, whereas the ethical type chooses the highest effort level ($e_2 = 1/a$). Therefore, voter i votes for the incumbent if and only if

$$(5) \Leftrightarrow p_i + \zeta_i + \eta \geq \frac{1}{2} \Leftrightarrow \zeta_i \geq \frac{1}{2} - \eta - p_i.$$

Probability of winning: By aggregating each voter's decision, we derive the incumbent's winning probability. Let the expected value of p_i among the electorate given (m, y_1) be $\mathbb{E}[p_i \mid m, y_1]$. The winning probability of the incumbent given (m, y_1) is

$$\Pr\left(\underbrace{\frac{1}{2} + \gamma\left(\mathbb{E}[p_i \mid m, y_1] - \frac{1}{2} + \eta\right)}_{\text{\# of the incumbent's votes given } \eta} \geq \frac{1}{2}\right) = \frac{1}{2} + \psi\left(\mathbb{E}[p_i \mid m, y_1] - \frac{1}{2}\right).$$

Furthermore, y_1 is a stochastic variable whose probability of taking the value one is ae_1. Therefore, the expected reelection probability of the incumbent given (m, e_1) is

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

where $\mathbb{E}[p_i \mid m, e_1] = ae_1\mathbb{E}[p_i \mid m, y_1 = 1] + (1 - ae_1)\mathbb{E}[p_i \mid m, y_1 = 0]$.

Hereafter, we assume the following inequality, which guarantees that the opportunistic type never exerts maximum effort for any δ:

Assumption 1. *$\psi a^2 b\Delta < 1$ holds.*

3.4 Confirmation Bias

Finally, we model voters' confirmation bias based on the formulation by Rabin and Schrag (1999). In response to the public signal, the incumbent's message and performance, voters update their beliefs on two uncertain states: the value of ω and the incumbent's type.

Learning the value of ω: Each voter receives a public signal s. Let voter i's subjective belief about the probability of ω being one just before the arrival of signal s be $\alpha_i \in \{\underline{\alpha}, \bar{\alpha}\}$. To model confirmation bias, we suppose that voters may misinterpret a signal contradicting the current belief. In particular, voter i perceives that the received signal is $\tilde{s}_i \in \{0, 1\}$. We allow $\tilde{s}_i \neq s$ to be the case when s contradicts his prior. The detailed assumption is as follows. Note that we do not consider the case where $\alpha_i = 0.5$ because $0 < \underline{\alpha} < 0.5 < \bar{\alpha}$ is assumed.

Assumption 2. *Voter i updates the belief about ω as follows:*

(i). Suppose that $\alpha_i > 0.5$. When $s = 1$, $\tilde{s}_i = 1$. When $s = 0$, $\tilde{s}_i = 1$ with probability q_i and $\tilde{s}_i = 0$ with the remaining probability.

(ii). Suppose that $\alpha_i < 0.5$. When $s = 0$, $\tilde{s}_i = 0$. When $s = 1$, $\tilde{s}_i = 0$ with probability q_i and $\tilde{s}_i = 1$ with the remaining probability.

(iii). Voter i does not know the possibility of misperception. That is, he updates the belief on the value of ω in a Bayesian manner as if $s = \tilde{s}_i$. The updated belief is denoted by $\tilde{\alpha}_i(\tilde{s}_i)$.

Suppose that voter i believes that state 1 is more likely to be true. Since signal 1 is consistent with his prior belief, he does not misperceive this signal. However, when he receives signal 0, he misperceives that he received the belief-consonant signal (i.e., signal 1) with probability q_i (see (i)).[14] The updating when he thinks that state 0 is more likely to be true is defined symmetrically (see (ii)). When $q_i = 0$, the situation is reduced to be Bayesian updating; thus, the value of $q_i \in [0, 1]$ represents the degree of confirmation bias. Lastly, each voter is naive in the sense that he does not notice that he is suffering from confirmation bias, which is (iii). As s is a perfect signal, the learning about ω ceases after the public signal.

[14]Hence, his initial opinion may be reinforced despite the information suggesting the opposite. This is consistent with experimental findings: people's beliefs could diverge after observing the same piece of information (e.g., Lord, Ross and Lepper, 1979; Fryer, Harms and Jackson, 2019); and correcting misinformation may increase misperceptions (e.g., Nyhan and Reifler, 2010) (to be fair, the results for the latter are mixed (e.g., Wood and Porter, 2019; Guess and Coppock, 2020).) The mechanism is as follows. First, corrective information is typically reported as news pitting on two sides of an argument. In this situation, agents may receive only a belief-consonant argument included in the news (e.g., Fryer, Harms and Jackson, 2019). Second, when receiving information that contradicts their opinion, agents may generate counter-arguments, which reinforces their initial opinion (e.g., Taber and Lodge, 2006; Taber, Cann and Kucsova, 2009). Note that this property is not critical for the present study. Even if we consider an alternative model that voters do not reinforce but just keep the prior by ignoring a belief inconsonant signal, our main results hold (see Section 6).

Learning the incumbent's type: Voters face two signals about the incumbent's type sequentially: the incumbent's message, m, and the incumbent's performance, y_1.

We start with the timing at which voters observe the incumbent's message. The prior probability of the incumbent being ethical is half. When the prior is half, voters reason that both possibilities are equally likely to be true. As the prior belief is neutral, they would never misperceive the message (Rabin and Schrag, 1999). Therefore, we assume the following:[15]

Assumption 3. *As the prior belief on the incumbent's type is neutral, voter i never misperceives m. Based on* $\tilde{\alpha}_i(\tilde{s}_i)$ *and the equilibrium strategy, he updates the probability of the incumbent being ethical in a Bayesian manner if it is possible. The updated belief is denoted by* $p_i^{int}(m)$.

After the prior of the incumbent's being ethical is updated to p_i^{int}, voters observe the incumbent's performance and further updates their belief. p_i^{int} is no longer half because of learning based on m. Hence, voter i may misperceive the incumbent's performance due to confirmation bias. Let the performance perceived by voter i be $\tilde{y}_{1i}$.

Assumption 4. *After observing* y_1, *voter i updates the belief on the incumbent's type as follows:*

(i). Suppose that $p_i^{int} > 0.5$. *When* $y_1 = 1$, $\tilde{y}_{1i} = 1$. *When* $y_1 = 0$, $\tilde{y}_{1i} = 1$ *with probability* q_i *and* $\tilde{y}_{1i} = 0$ *with the remaining probability.*

(ii). Suppose that $p_i^{int} < 0.5$. *When* $y_1 = 0$, $\tilde{y}_{1i} = 0$. *When* $y_1 = 1$, $\tilde{y}_{1i} = 0$ *with probability* q_i *and* $\tilde{y}_{1i} = 1$ *with the remaining probability.*

(iii). Voter i does not know the possibility of misperception. Based on $\tilde{\alpha}(\tilde{s}_i)$, $p_i^{int}(m)$, *and the equilibrium strategy, he updates the belief in a Bayesian manner as if* $y_1 = \tilde{y}_{1i}$, *as long as it is possible. The updated belief is denoted by* $p_i(m, \tilde{y}_{1i})$.

$y_1 = 1$ serves as an imperfect signal of the incumbent being ethical because the ethical type exerts the maximum effort and always achieves high performance. Hence, voter i may misperceive low performance as high performance if $p_i^{int} > 0.5$ and vice versa if $p_i^{int} < 0.5$.

We have two remarks. First, the perceived signal is randomly drawn across time and individuals. Second, the value of q_i for learning ω and that for learning the incumbent's type are common within each individual. The latter assumption will be relaxed in Section 6.

[15] Though the learning process about ω ends before this timing, m serves as a signal about ω as well as the incumbent's type. Hence, the voter might misperceive the incumbent's message to confirm his belief on ω, which influences learning about the incumbent's type. We do not consider such a possibility by assuming that he updates the belief about ω and the incumbent's type *separately*; that is, he separately updates the marginal probability instead of the joint probability because updating a joint distribution is cognitively taxing (Loh and Phelan, 2019). Thus, the voter does not misperceive m in learning the incumbent's type even if he misperceives it in learning ω.

Heterogeneity: We introduce individual differences in the degree of confirmation bias by supposing that q_i follows a distribution F (either continuous or discrete), whose variance is given by σ^2. F can be any distribution. Note that $q \in [0, 1]$, and thus the upper bound of the variance is $\mathbb{E}[q](1 - \mathbb{E}[q])$. A higher σ^2 represents larger heterogeneity in the degree of confirmation bias across voters. F is common knowledge across all players.

3.5 Timing of the Game and Equilibrium Concept

The timing of the game is given as follows:

1. Voters observe a public signal, s, and the incumbent sends a message, m.
2. The incumbent exerts an effort, e_1, which is unobservable to voters.
3. After observing the performance level, y_1, voters vote for the incumbent or the challenger.
4. The elected politician exerts an effort, e_2, and the performance level, y_2, is realized.

The equilibrium concept is a variant of pure-strategy perfect Bayesian equilibrium. Specifically, the opportunistic-type's strategy and voters' belief system constitute an equilibrium if (i) the opportunistic-type's equilibrium strategy maximizes her payoff given voters' behaviors induced by their equilibrium belief system and (ii) given the opportunistic-type's equilibrium strategy, voters' equilibrium belief system is consistent with Assumptions 2-4. Assumptions 2-4 have no restriction on off-path beliefs. To focus on reasonable ones, we assume that each voter believes that the incumbent is opportunistic if he thinks that an off-path is observed.

4 Equilibrium

4.1 Belief Polarization after the Public Signal

At the beginning of the game, voters receive a public signal $s = 1$ because $\omega = 1$; however, some voters misperceive the signal, which makes public opinion concerning the value of ω polarized.

First, we consider the κ fraction of voters who initially believed that state 1 is more likely to be true. Signal 1 confirms their initial hypothesis so they perceive signal 1 literally (i.e., $\tilde{s}_i = 1$). Hence, for all of them, the posterior probability of ω being one, $\tilde{\alpha}_i$, is one. Next, we consider $1 - \kappa$ fraction of voters who initially believed that state 0 is more likely to be true. Signal 1 contradicts their initial hypothesis, so they may misperceive signal 1 as signal 0. Particularly, voter i misperceives the signal (i.e., $\tilde{s}_i = 0$) and believes that the truth is $\omega = 0$ (i.e., $\tilde{\alpha}_i = 0$) with probability q_i, whereas he perceives signal 1 (i.e., $\tilde{s}_i = 1$) and learns the truth (i.e., $\tilde{\alpha}_i = 1$) with probability $1 - q_i$. Hence, among the $1 - \kappa$ fraction of voters, $\mathbb{E}[q]$ fraction believes the misinformation, whereas $1 - \mathbb{E}[q]$ fraction believes the truth.

Consequently, despite voters observing the same public signal, their beliefs diverge more due to confirmation bias (Rabin and Schrag, 1999). This belief polarization after observing the same piece of information looks odds at first glance, but it is indeed consistent with empirical findings. In the seminal experiment by Lord, Ross and Lepper (1979), participants were assigned research summaries about the deterrent efficacy of the death penalty. They found that subjects' beliefs were polarized despite reading the same piece of information.[16]

In summary, $\kappa + (1-\kappa)(1-\mathbb{E}[q])$ fraction of voters are *informed* in that they firmly believe the truth (i.e., $\tilde{\alpha}_i = 1$). On the other hand, the remaining fraction of voters are *misinformed* in that they firmly believe the misinformation (i.e., $\tilde{\alpha}_i = 0$). If the majority of voters are misinformed, it should not be surprising that the incumbent supports the misinformation. To analyze the interesting case where only a minority of voters believe in it, we assume the following:

Assumption 5. $\kappa + (1 - \kappa)(1 - \mathbb{E}[q]) > \frac{1}{2}$ *holds.*

4.2 Incumbent's Message and Political Support Base

After the public signal, the incumbent supports or denies 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 δ when she is strategic and the state is ω be $m^*(\delta, \omega)$ and the expected value of $m^*(\delta, \omega)$ over δ be $\mathbb{E}[m^*(\delta, \omega)] := \int_0^{\Delta} \frac{m^*(\delta,\omega)}{\Delta} d\delta$. Without loss of generality, we assume symmetric strategies such that whether to confirm the truth is independent of ω i.e., $m^*(\delta, 1) = 1 - m^*(\delta, 0)$. In addition, with probability ε, the opportunistic type non-strategically denies the truth. Hence, the opportunistic type confirms the truth with probability $(1 - \varepsilon)\mathbb{E}[m^*(\delta, \omega)]$.

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. Hence, they literally perceive

[16] Subsequent works support their finding (e.g., Taber and Lodge, 2006; Taber, Cann and Kucsova, 2009; Fryer, Harms and Jackson, 2019). Note that in a rich setting, such as with multidimensional uncertainty, Bayesian learners may experience belief polarization (e.g., Andreoni and Mylovanov, 2012).

the incumbent's message.[17] However, as informed and misinformed voters believe that a different state is true, their evaluation of the incumbent's message differs. 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 after the incumbent's performance. The belief is derived as follows:

Lemma 1. *(i). Informed voters evaluate the incumbent as follows:*

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

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

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

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

Second, we consider misinformed voters who falsely believe that $\omega = 0$. For them, $m = 1$ is regarded as denying the truth, although its actual meaning is to claim the truth. Therefore, the incumbent's reputation becomes zero among them. On the contrary, $m = 0$ is regarded as confirming the truth. Without noticing the actual meaning, they upwardly update the reputation.

Once reputation becomes zero, it will not be restored; thus, they vote for the incumbent only when $\zeta_i + \eta > 0.5$.[18] Hence, whether to confirm the truth determines who is the incumbent's political support base. We say that the incumbent's political support base consists of informed (resp. misinformed) voters if $p_I^{int} > 0.5$ and $p_M^{int} = 0.5$ (resp. $p_M^{int} > 0.5$ and $p_I^{int} = 0.5$). If the incumbent confirms the truth, her political support base consists of informed voters, while it consists of misinformed voters if she denies the truth. From Assumption 5, the majority of

[17]In practice, a fraction of voters initially have either strong anti- or pro-incumbent beliefs. We ignore them because their voting choice would not change in response to the incumbent's message and electoral results are determined by swing voter. Indeed, Swire-Thompson et al. (2020) experimentally show that correcting misinformation by a politician does not change the evaluation among subjects having pro- or anti-sentiment toward the politician.

[18]While voters outside the political support base are unlikely to vote for the incumbent and their voting probability is no longer dependent upon y_1, the probability of such voters voting for the incumbent is not zero because of probabilistic voting. Provided that ζ_i can be interpreted as voter i's partisanship, this captures the reality that voters with strong partisanship vote for the incumbent even if the reputation is quite low.

voters are informed. Hence, in terms of maximizing the political support base, the incumbent should not deny the truth. Later, we prove that the opportunistic type may strategically deny the truth despite this electoral cost.

4.3 Reelection Probability

As we have shown in Section 3.3, the reelection probability given (m, e_1) is

$$P^*(m, e_1) = \frac{1}{2} + \psi\left(\mathbb{E}[p_i \mid m, e_1] - \frac{1}{2}\right), \tag{6}$$

implying that it is increasing in the incumbent's expected average reputation, $\mathbb{E}[p_i \mid m, e_1]$. To derive the value of $\mathbb{E}[p_i \mid m, e_1]$, we first explore how the incumbent's reputation is updated among each voter. If the voter is outside the incumbent's political support base, the incumbent's reputation for him is zero. If the voter is in the incumbent's political support base, the incumbent's reputation for him depends on whether he perceives that the incumbent's performance is high. Let the perceived performance for voter i be $\tilde{y}_{1i}$. On the one hand, if $\tilde{y}_{1i} = 0$, the incumbent's reputation becomes zero for him. On the other hand, if $\tilde{y}_{1i} = 1$, the incumbent's reputation is updated upwardly. In summary, voter i positively evaluates the incumbent if and only if he is in the political support base and perceives that the incumbent's performance is high.

Based on this, the incumbent's expected average reputation given (m, e) is obtained:

$$\mathbb{E}[p_i \mid 1, e_1] = \underbrace{[\kappa + (1-\kappa)(1-\mathbb{E}[q])]}_{\text{Size of } I} \underbrace{[ae_1 + \mathbb{E}[q_i \mid I](1-ae_1)]}_{\text{Expected prob. of } \tilde{y}_{1i}=1 \text{ among } I} \underbrace{\frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}}_{\text{Reputation for } I \text{ with } \tilde{y}_{1i}=1};$$

$$\mathbb{E}[p_i \mid 0, e_1] = \underbrace{(1-\kappa)\mathbb{E}[q]}_{\text{Size of } M} \underbrace{[ae_1 + \mathbb{E}[q_i \mid M](1-ae_1)]}_{\text{Expected prob. of } \tilde{y}_{1i}=1 \text{ among } M} \underbrace{\frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}}_{\text{Reputation for } M \text{ with } \tilde{y}_{1i}=1},$$

where $\mathbb{E}[q_i \mid I]$ (resp. $\mathbb{E}[q_i \mid M]$) is the average value of q_i (i.e., the average degree of confirmation bias) among informed (resp. misinformed) voters, and $\mathbb{E}[y_o^* \mid m]$ is the equilibrium probability of the opportunistic type's performance being high.

To understand these formulas, first, consider the case where the incumbent supports the truth. In this case, the political support base consists of informed voters. Thus, the average reputation is given by (the size of informed voters) × (the expected average reputation among informed voters). Informed voter i receives $\tilde{y}_{1i} = 1$ either when the incumbent's performance is high or when low performance is misperceived as high. Therefore, the reputation is positive with probability $ae_1 + q_i(1 - ae_1)$. Now, the value of q_i is heterogeneous across voters, and its expected value among informed voters is $\mathbb{E}[q_i \mid I]$. Therefore, the expected fraction of informed voters perceiving the incumbent's high performance is given by $ae_1 + \mathbb{E}[q_i \mid I](1 - ae_1)$. Combining them yields the above formula for $\mathbb{E}[p_i \mid 1, e_1]$.

Next, suppose that the incumbent denies the truth. In this case, the political support base is misinformed voters. Thus, what matters is the number of misinformed voters and the expected value of q_i among misinformed voters. Hence, we obtain the above formula of $\mathbb{E}[p_i \mid 0, e_1]$.

An important observation is that the incumbent's popularity depends not only on the size of the political support base but also on the expected degree of confirmation bias among the political support base. If performance is low, the incumbent's reputation should be severely undermined. However, voters among the political support base may misperceive low performance as high because of confirmation bias, which increases the incumbent's reputation. Therefore, the expected degree of confirmation bias among the political support base matters.

Substituting these formulas into (6) yields the incumbent's reelection probability:

Lemma 2. *Given (m, e), the incumbent's reelection probability is:*

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

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

Hence, the following holds:

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

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

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

(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 (Besley, 2006).

(ii) shows that $P^*(m, e)$ is increasing in the average confirmation bias among the political support base. Voters in the support base may misperceive low performance as high because they have a belief that the incumbent is likely to be ethical, and a higher confirmation bias increases this misperception probability. Thus, the higher the average confirmation bias among the support base, the less likely the incumbent is to be punished for low performance. Consequently, the higher confirmation bias among the support base is better for the incumbent.

Lastly, as in (iii), the marginal benefit of effort is decreasing in the average confirmation bias among the political support base. The higher the average bias among the political support base, the less likely it is that low performance is to be punished. Hence, the marginal benefit of exerting an effort is reduced. (ii) and (iii) are in line with the finding by Lockwood (2017).[19]

4.4 Average Confirmation Bias in the Political Support Base

As Lemma 2 (ii) shows, a key determinant of the reelection probability is the average confirmation bias among the political support base; that is, $\mathbb{E}[q_i \mid I]$ when supporting the truth and $\mathbb{E}[q_i \mid M]$ when denying the truth. A simple calculation yields an explicit formula for these values:

$$\mathbb{E}[q_i \mid I] = \frac{\kappa\mathbb{E}[q] + (1-\kappa)[\mathbb{E}[q] - \mathbb{E}[q^2]]}{\kappa + (1-\kappa)[1 - \mathbb{E}[q]]}; \quad \mathbb{E}[q_i \mid M] = \frac{\mathbb{E}[q^2]}{\mathbb{E}[q]}.$$

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

Lemma 3. $\mathbb{E}[q_i \mid I] \leq \mathbb{E}[q_i \mid M]$*, where the equality holds if and only if* $\sigma^2 = 0$*. Furthermore,* $\mathbb{E}[q_i \mid I]$ *is decreasing in* σ^2*, whereas* $\mathbb{E}[q_i \mid M]$ *is increasing in* σ^2*.*

First, suppose that the degree of confirmation bias is homogeneous (i.e., $\sigma^2 = 0$). 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 among informed voters is strictly lower than that among misinformed voters when bias is heterogeneous (i.e., $\sigma^2 > 0$). Therefore, denying the truth means that voters with high confirmation bias become the political support base, which will be key to understanding our main result. Furthermore, this difference in the average degree of bias is increasing in σ^2.

The intuition for the heterogeneous bias case is as follows. As voters receive a public signal about ω, they should all be informed if they are rational. However, misinformed voters exist because some voters misperceive the public signal owing to confirmation bias. As the probability of misperception increases with q_i, the average bias is higher among misinformed voters.

4.5 When the Incumbent Denies the Truth

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

[19]It is notable that (iii) holds under the setting à la Rabin and Schrag (1999), but may not hold in a different type of motivated reasoning (Little, Schnakenberg and Turner, 2022).

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 σ^2 is not so high). In this case, the incumbent never strategically denies the truth.

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

To see this, suppose an extreme case wherein $\sigma^2 = 0$. Lemma 3 indicates that $\sigma^2 = 0$ implies $\mathbb{E}[q_i \mid I] = \mathbb{E}[q_i \mid M]$; thus, from Lemma 2, $P^*(1, e) > P^*(0, e)$ holds under Assumption 5. In words, denying the truth shrinks the political support base. Since the smaller support base reduces the number of obtained votes, the incumbent should not deny the truth. This result remains as long as σ^2 is not very large.

Case with high heterogeneity: When the degree of confirmation bias is sufficiently heterogeneous across voters, this outcome no longer holds:

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

(i). The equilibrium is characterized by a unique $\bar{\delta} \in (0, \Delta]$. *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* σ^2.

When the degree of confirmation bias is sufficiently heterogeneous across voters, the opportunistic type may strategically deny the truth. In particular, the opportunistic type 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 result can be understood as follows. 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, $\mathbb{E}[q_i \mid M] > \mathbb{E}[q_i \mid I]$ holds (see Lemma 3). As a result, targeting misinformed voters as a political support base enables the incumbent to weaken electoral punishment after the low performance (Lemma 2 (ii)).

Therefore, the incumbent faces both pros and cons of denying the truth. On the one hand, the electoral cost is that her political support base shrinks. On the other hand, its electoral 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 is expected to be low; thus, for them, the benefit dominates the cost when the degree of confirmation bias is sufficiently heterogeneous. As such, low-competent politicians deny the truth supported by the majority of voters to capture voters with high confirmation bias as the political support base.

Before closing this subsection, we summarize how the threshold value, $\bar{\sigma}^2$, is determined:

Corollary 1. *$\bar{\sigma}^2$ is increasing in κ.*[20] *It is also 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 fraction of voters who initially believed the truth, κ. A larger κ means that a larger number of voters support the truth. Hence, it lowers the incumbent's incentive to deny the truth.

Another determinant is the average degree of confirmation bias, $\mathbb{E}[q]$. The higher confirmation bias has three effects. First, it increases the average bias among informed voters. Second, the same holds for misinformed voters, too. Third, it increases the number of misinformed voters. The first lowers the incumbent's incentive to deny the truth, whereas the second and third have the opposite effect. On the one hand, when κ 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 κ is low, the fraction of misinformed voters is large; thus a higher $\mathbb{E}[q]$ encourages the incumbent's misinformation.

4.6 Effect on the Incumbent's Effort for Public Goods Provision

We have shown that denying the truth weakens electoral punishment. Hence, it might influence the incumbent's effort level for public goods provision. Regarding this issue, we obtain the following proposition. Let the equilibrium effort level of the opportunistic type if she is strategic be $e^{**}(\delta) := e^*(\delta, m^*(\delta, 1))$, where $e^*(\delta, m)$ is her equilibrium effort given m.

Proposition 3. *(i). Suppose that $\sigma^2 \leq \bar{\sigma}^2$. Then, $e^{**}(\delta)$ is increasing in δ.*

(ii). Suppose that $\sigma^2 > \bar{\sigma}^2$. Then, $e^{**}(\delta)$ is increasing in δ and $\lim_{\delta \nearrow \bar{\delta}} e^{**}(\delta) < \lim_{\delta \searrow \bar{\delta}} e^{**}(\delta)$.

[20] The upper bound of σ^2 is $\mathbb{E}[q](1 - \mathbb{E}[q])$. When $\kappa \geq 1/2$, $\bar{\sigma}^2 \geq \mathbb{E}[q](1 - \mathbb{E}[q])$ i.e., $\sigma^2 \geq \bar{\sigma}^2$ never holds.

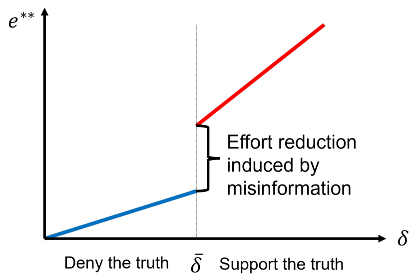

Figure 2: Effort and Misinformation

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

Second, suppose that σ^2 exceeds $\bar{\sigma}^2$; that is, the opportunistic type never strategically 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 is that the effort level drops discontinuously when δ falls below the threshold, despite the cost of effort being close to each other around the threshold. This discontinuous drop identifies the negative effect of strategic misinformation.[21] When the incumbent denies the truth, her political support base consists of misinformed voters who are less likely to punish low performance. As a result, the marginal benefit of effort drops discontinuously once the incumbent denies the truth, which reduces the effort level (Lemma 2 (iii)).

This finding reveals the hidden welfare costs associated with misinformation. It is believed that politicians' denying the truth induces citizens to hold wrong beliefs. For example, President Bolsonaro's misinformation about COVID-19 in Brazil was criticized because it forced Brazilians to experience severe COVID-19 situations (Ricard and Medeiros, 2020). Our result uncovers another hidden cost of misinformation. Low-competent politicians deny the truth to weaken the electoral punishment for low performance. Therefore, misinformation has a negative spill-over effect: it lowers politicians' effort level for providing public goods, reducing voters' welfare. This hidden cost should be considered when evaluating the welfare implications.

Effect of heterogeneous confirmation bias: Given the above findings, it might be expected that larger heterogeneity in confirmation bias reduces the effort level. Interestingly, this view may not be true. To see this, let $e^{**}(\delta)$ given σ^2 be $e^{**}(\delta \mid \sigma^2)$ and $\bar{\delta}$ given σ^2 be $\bar{\delta}(\sigma^2)$.

[21] For simplicity, we have assumed that the incumbent claims the truth in an indifferent case so that the incumbent does so when $\delta = \bar{\delta}$. It is notable that this tie-breaking rule does not affect the discontinuity result.

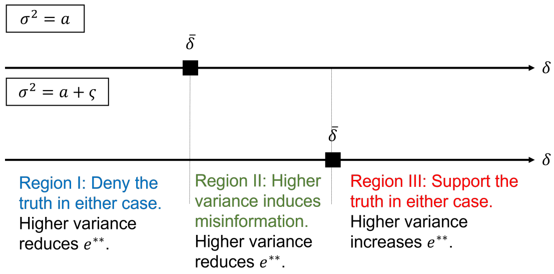

Figure 3: Larger Heterogeneity and Effort

Proposition 4. *Fix $\mathbb{E}[q]$ and suppose that σ^2 changes from v to $v + \varsigma$, where ς is a sufficiently small, but positive value.*

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

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

First, when the initial value of σ^2 is small, a marginal change of σ^2 increases the opportunistic type's effort level as in (i). The mechanism is as follows. As larger heterogeneity reduces the average confirmation bias among informed voters (Lemma 3), it stimulates the effort level of politicians whose political support base consists of informed voters. Given the low heterogeneity, every politician supports the truth so that her political support base consists of informed voters. Therefore, a marginal increase in σ^2 has a positive effect on every politician's effort level.

Second, suppose that the initial value of σ^2 is large ((ii) of the above proposition). In this case, a marginal change of σ^2 reduces the opportunistic type's effort level if and only if δ falls below a threshold value. The mechanism is as follows. Consider low-competent politicians who deny the truth in either value of σ^2 (i.e., Region I in Figure 3). Greater heterogeneity increases the average confirmation bias among their political support base, misinformed voters (Lemma 3). Therefore, larger heterogeneity discourages effort. Next, consider politicians in Region II. They initially supported the truth, but started to deny the truth, which reduces the level of effort. Lastly, consider politicians in Region III. They support the truth in either case, so their political support base consists of informed voters. As the larger heterogeneity reduces the average confirmation bias among informed voters (Lemma 3), the effort level is stimulated. In summary, larger heterogeneity induces low-competent politicians to strategically deny the truth and reduce their effort, but it also encourages high-competent politicians' effort.

5 Application: Effect of Social Media Platforms

There is widespread concern that social media platforms may facilitate the spread of fake news and create echo chambers by allowing users to selectively expose themselves to information that reinforces their preexisting views (e.g., Del Vicario et al., 2016; Allcott and Gentzkow, 2017; Levy, 2021; Acemoglu, Ozdaglar and Siderius, 2021). Our framework can be utilized to analyze the impact of these social media platforms on political misinformation.

Setting: For this purpose, we provide a micro-foundation of our learning process. Suppose that voter i confronts at most two pieces of information about s. The first one is the true signal s. The second one is the belief-consonant fake signal $f \in \{0, 1\}$, which takes one if $\alpha_i > 0.5$ and zero if $\alpha_i < 0.5$. The true signal is evident so that rational agents distinguish between s and f. However, irrational agents may misperceive f as s (i.e., $\tilde{s}_i = f$) to confirm their prior belief. The misperception probability is q_{ib} and voters are unaware of misperception.

The availability of a belief-consonant fake signal depends on the information environment. When only the mainstream media is available, voters are unlikely to be subjected to fake news. On the contrary, recent developments of social media platforms have enabled people to expose themselves to pro-attitudinal fake news. To describe this, we assume that voter i receives f in addition to s with probability $\theta \in (0, 1)$. Voter i's learning about ω is summarized as follows:

- With probability $1 - \theta$, voter i receives only the true signal so that $\tilde{s}_i = s$.
- With probability $\theta(1-q_{ib})$, voter i receives both the true signal and the belief-consonant fake signal, but correctly understands the true signal so that $\tilde{s}_i = s$.
- With probability θq_{ib}, voter i receives both the true signal and the belief-consonant fake signal, and misperceives the true signal so that $\tilde{s}_i = f$.

In summary, the probability of misperception, q_i, is θq_{ib}. As for learning about the incumbent's type, we consider the same setup. This setting provides a foundation for the learning process in Assumption 2 and 4. A notable feature is that q_i is a joint product of the psychological factor, q_{ib}, and the information environment, θ. Though we have allowed any distribution in the main analysis, we assume that q_{ib} follows $U[0, B]$; thus, q_i follows $U[0, \theta B]$, where $\theta B < 1$.

Effect of social media platforms: The development of social media platforms increases the value of θ through the availability of the belief-consonant fake news. This has two effects on q_i. First, $\mathbb{E}[q_i]$ increases; thus, the number of misinformed voters rises as well. Second, the variance, σ^2, also increases because it is given by $\frac{\theta^2 B^2}{12}$.

The first effect through an increase in $\mathbb{E}[q_i]$ corresponds to a widely shared concern about fake news and echo chambers (e.g., Del Vicario et al., 2016). Larger $\mathbb{E}[q_i]$ increases the number of misinformed voters. Hence, social media platforms cause a large fraction of voters misinformed. However, its effect on the political supply of misinformation is ambiguous as depicted in Corollary 1: it fosters the incumbent to deny the truth only if κ is sufficiently low.

In addition, our model identifies a new effect through a larger σ^2. Proposition 2 indicates that a larger σ^2 fosters politicians' use of misinformation. Hence, the emergence of social media platforms encourages politicians to deny the truth through the larger heterogeneity of q_{ib}. This latter effect has been ignored in the literature, but should be considered.

The total effect of higher θ can be obtained as follows:

Proposition 5.

$$\sigma^2 > \bar{\sigma}^2 \iff \theta > \frac{3}{4(1-\kappa)B}.$$

Therefore, low-competent politicians deny the truth if and only if θ is sufficiently large. As discussed above, this should not be seen as an outcome of the larger average probability of misperception because a larger average probability of misperception may discourage the incumbent to deny the truth. This proposition highlights that the effect through the larger variance is a dominant force. Consequently, a larger θ always encourages the incumbent to deny the truth, which further could reduce the low-competent incumbent's effort level.

Scholars have recently argued that social media platforms should conduct fact-checking (Allcott, Gentzkow and Yu, 2019) and implement algorithms to mitigate echo chamber effects (Levy, 2021). The typical reasoning behind this argument is that such efforts reduce the value of θ, thereby decreasing the proportion of misinformed citizens. Although we agree on its importance, Proposition 5 provides another rationale for the importance of such attempts. Without them, individuals with high q_{ib} are vulnerable to misinformation, whereas those with low q_{ib} are not as susceptible to confirmation bias. Consequently, the variance of q_i among citizens increases, enabling politicians to deny the truth and undermine electoral accountability. This reduces low-competent politicians' effort level. Therefore, when the incumbent is likely to be low-competent, fake news and echo chambers should be regulated because such attempt reduces the heterogeneity of confirmation bias and prevents politicians from weakening electoral accountability.[22] Our results highlight this overlooked role.

[22]As in Proposition 4, the larger heterogeneity fosters the high-competent politician's effort. Hence, reducing the heterogeneity is beneficial for voters only when the incumbent is likely to be low-competent.

6 Discussion

6.1 Implications for Empirical Research

Whether denying the truth fosters the chance of electoral success is an important empirical question (e.g., Barrera et al., 2020; Malzahn and Hall, 2023). 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 addressed 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. Without accounting for this, we may obtain a negative correlation between denying the truth and electoral outcome even if denying the truth is beneficial. In addition, finding a quasi-experimental situation is still unsatisfactory. The causal effect of denying the truth given δ is $P^*(0, e^*(\delta, 0)) - P^*(1, e^*(\delta, 1))$. This is heterogeneous depending on the competence level, δ. In particular, it can be easily shown that it is positive for small δ, but negative for large δ when $\sigma^2 > \bar{\sigma}^2$ (see the Online Appendix). 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 involves conducting experimental research. A straightforward approach would be to conduct a conjoint survey experiment wherein each subject is assigned to a politician with various attributes and asked whether they support the politician (Bansak et al., 2021). By including whether the politician claims misinformation as an attribute, we could identify the electoral effect of denying the truth. However, this approach should be improved. The electoral benefit of denying the truth is that it weakens electoral punishment when low performance is realized. Hence, it is not appropriate to measure voters' support for a politician just after the politician denies the truth. We need to measure voters' support for the politician when her low performance is realized sometime after she denied the truth.[23]

[23]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. Indeed, the literature of confirmation bias has shown that the order of information arrivals matters in contrast to Bayesian updating (Rabin and Schrag, 1999; Fryer, Harms and Jackson, 2019).

6.2 Extensions

The details of Sections 6.2 and 6.3 are available in the Online Appendix.

Challenger's message: Thus far, we have assumed that only the incumbent sends a message, m. However, the challenger may also support or deny the truth. We show that the challenger has no incentive to deny the truth. The mechanism is as follows. The incumbent denies the truth in order to weaken electoral punishment for low performance. On the contrary, the challenger exerts no effort in the provision of public goods; thus 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 the size of her political support base. Therefore, denying the truth is not optimal for her. It should be noted that this does not imply that the challenger is always less likely to deny the truth, because there exists a selection issue omitted from the model. In practice, some challengers may believe the false argument and deny the truth for non-strategic reasons, whereas this occurrence is unlikely for the incumbent because such politicians lose an election and cannot be the incumbent.

Correlation of confirmation bias across issues: The degree of confirmation bias when learning ω and that when learning the incumbent's type may differ even within individuals. To consider this, suppose that these two are the same with probability λ, but independent with the remaining probability. It is shown that $\bar{\sigma}^2$ is increasing in λ. To see the mechanism, suppose an extreme case wherein the correlation is zero. Then, voters believing the false argument about ω have a high confirmation bias when learning the truth on this issue, but their confirmation bias when learning the incumbent's type is not necessarily low. Hence, attracting voters who believe the false argument about ω does not help weaken electoral punishment. The effect of the heterogeneity of confirmation bias depends on whether it is across issues or across individuals.

6.3 Discussion on Modeling Assumptions

Repeated elections with term-limits: We can easily extend the model to an infinitely repeated elections model with two term-limits (e.g., Banks and Sundaram, 1998; Smart and Sturm, 2013; Kartik and Van Weelden, 2019*b*). Our results remain robust under this more realistic setting, provided that we focus on Markov perfect equilibria.

Strategic incentive of the ethical type: Thus far, we have assumed that the ethical type non-strategically tells the truth and exerts the maximum effort. Consider an alternative setting that the ethical type behaves to maximize voters' welfare. It is shown that similar results hold; that is, the assumption that the ethical type is non-strategic is innocuous.

Probability of misperception: Following Rabin and Schrag (1999), we have assumed that the probability of misperception is independent of the strength of the prior belief. However, in practice, the probability of misperception may be increasing in the extremeness of the prior. By exploring this alternative setting, we demonstrate that the incumbent has no strategic incentive to deny the truth if and only if the degree of heterogeneity of confirmation bias sufficiently small. In this sense, the key property remains valid under this alternative setting.

The incumbent's prior reputation: We have assumed that the prior of the incumbent being ethical is half; thus, voters update their belief in an unbiased way when observing the incumbent's message. As an extension, we allow the prior not to be half in the alternative setting of confirmation bias introduced above. As long as the prior on the incumbent's type is moderate and the prior on ω is extreme, the incumbent has no incentive to deny the truth if and only if the degree of heterogeneity of confirmation bias is sufficiently small. Furthermore, under a different scenario, for any prior on the incumbent's type, the same results could hold.

Weaker version of confirmation bias: In line with Rabin and Schrag (1999), we have assumed that with a positive probability, voters misinterpret a belief-inconsonant signal as the opposite signal. Hence, voters' misperceptions may be reinforced after they observe the correct information. Another possibility is a weaker version of confirmation bias: with a positive probability, voters simply ignore a belief-inconsonant signal and do not update the prior. This setting complicates the analysis, but the mechanism remains the same. As long as $\underline{\alpha}$ is close to zero, the incumbent has no strategic incentive to deny the truth if and only if the degree of heterogeneity of confirmation bias is sufficiently small.

Policy choice: We have assumed that the incumbent determines the effort level for the provision of public goods. Another conflict of interest between politicians and voters is concerning policy choice rather than effort choice (i.e., they have different ideal policies). Even in this case, the incumbent has no strategic incentive to deny the truth if and only if the degree of heterogeneity of confirmation bias is sufficiently small.

7 Concluding Remarks

In the current political landscape, elected politicians sometimes support false claims. We examined this issue by extending a canonical two-period electoral accountability model. We showed that the low-competent incumbent has an electoral incentive to supply misinformation even if only a minority of voters believe in the misinformation. The key to understanding this paradox is the heterogeneity of confirmation bias, a psychological bias that has been identified in the literature as a determinant of people's persistent acceptance of misinformation. We showed that a low-competent politician denies the truth if and only if the degree of confirmation bias is sufficiently heterogeneous across voters. This suggests that a difference in the intra-country variance of confirmation bias may explain cross-country variations of politicians' use of misinformation.[24] Furthermore, our analysis showed that strategic misinformation enables low-competent politicians to weaken the fundamental role of an election as a device for disciplining politicians, which reduces their effort level. This negative side-effect on the performance of representative democracy should be considered.

To explore the impact on electoral accountability, we supposed that the incumbent politician exerts effort to produce public goods, a common interest issue for voters. On the contrary, issues such as distributive politics are characterized by divided interests across voters. In such a case, the incumbent may cater benefits to her political support base. Analyzing how claiming misinformation influences such divided issues is remaining for future research.

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.

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

[24]For example, the level of directional motivated reasoning is associated with the degree of political sophistication (eg., Taber and Lodge, 2006; Vegetti and Mancosu, 2020), which may imply that the variance of confirmation bias is correlated with that of political sophistication. If this is the case, our result would indicate that the variance of political sophistication is a determinant of politicians' use of misinformation.

Allcott, Hunt, Matthew Gentzkow, and Chuan Yu. 2019. "Trends in the diffusion of misinformation on social media." *Research & Politics*, 6(2): 1–8.

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.

Ardèvol-Abreu, Alberto, Homero Gil de Zuniga, and Elena Gámez. 2020. "The influence of conspiracy beliefs on conventional and unconventional forms of political participation: The mediating role of political efficacy." *British Journal of Social Psychology*, 59(2): 549–569.

Banks, Jeffrey S, and Rangarajan K Sundaram. 1998. "Optimal retention in agency problems." *Journal of Economic Theory*, 82(2): 293–323.

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.

Bräuninger, Thomas, and Nikolay Marinov. 2022. "Political elites and the "War on Truth"." *Journal of Public Economics*, 206: 104585.

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.

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.

Faia, Ester, Andreas Fuster, Vincenzo Pezone, and Basit Zafar. forthcoming. "Biases in information selection and processing: Survey evidence from the pandemic." *The Review of Economics and Statistics*.

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. 2023. "Reference points and democratic backsliding." *American Journal of Political Science*, 67(1): 71–88.

Grossman, Gene M, and Elhanan Helpman. 2023. "Electoral competition with fake news." *European Journal of Political Economy*, 77: 102315.

Guess, Andrew, and Alexander Coppock. 2020. "Does counter-attitudinal information cause backlash? Results from three large survey experiments." *British Journal of Political Science*, 50(4): 1497–1515.

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

Hohnsbehn, Jana-Maria, David F Urschler, and Iris K Schneider. 2022. "Torn but balanced: Trait ambivalence is negatively related to confirmation." *Personality and Individual Differences*, 196: 111736.

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.

Kartik, Navin, and Richard Van Weelden. 2019*a*. "Informative cheap talk in elections." *The Review of Economic Studies*, 86(2): 755–784.

Kartik, Navin, and Richard Van Weelden. 2019*b*. "Reputation effects and incumbency (dis) advantage." *Quarterly Journal of Political Science*, 14(2): 131–57.

Kartik, Navin, and R Preston McAfee. 2007. "Signaling character in electoral competition." *American Economic Review*, 97(3): 852–870.

Kocak, Korhan, and Özgür Kıbrıs. 2023. "Social media and press freedom." *British Journal of Political Science*, 53(1): 140–162.

Lasser, Jana, Segun Taofeek Aroyehun, Almog Simchon, Fabio Carrella, David Garcia, and Stephan Lewandowsky. 2022. "Social media sharing of low quality news sources by political elites." *PNAS Nexus*, 1(4): pgac186.

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

Levy, Ro'ee. 2021. "Social media, news consumption, and polarization: Evidence from a field experiment." *American Economic Review*, 111(3): 831–70.

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

Li, Anqi, and Lin Hu. 2023. "Electoral accountability and selection with personalized information aggregation." *Games and Economic Behavior*, 140: 296–315.

Li, Anqi, Davin Raiha, and Kenneth W Shotts. 2022. "Propaganda, alternative media, and accountability in fragile democracies." *The Journal of Politics*, 84(2): 1214–1219.

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.

Malzahn, Janet, and Andrew B. Hall. 2023. "Election-denying republican candidates underperformed in the 2022 midterms." *Working Paper*.

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.

Pirro, Andrea LP, and Paul Taggart. 2023. "Populists in power and conspiracy theories." *Party Politics*, 29(3): 413–423.

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).

Smart, Michael, and Daniel M Sturm. 2013. "Term limits and electoral accountability." *Journal of Public Economics*, 107: 93–102.

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).

Vegetti, Federico, and Moreno Mancosu. 2020. "The impact of political sophistication and motivated reasoning on misinformation." *Political Communication*, 37(5): 678–695.

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

Wood, Thomas, and Ethan Porter. 2019. "The elusive backfire effect: Mass attitudes' steadfast factual adherence." *Political Behavior*, 41: 135–163.

A Omitted Proofs

Proof of Lemma 2 Let informed (resp. misinformed) voter i's subjective probability of the incumbent being ethical given $(m, \tilde{y}_{1i})$ be $p_{Ii}(m, \tilde{y}_{1i})$ (resp. $p_{Mi}(m, \tilde{y}_{1i})$). To derive the value of p_i, we need to define the equilibrium probability of the opportunistic type's performance being high. With probability $1 - \varepsilon$, the opportunistic type is strategic and chooses the effort level maximizing her payoff, denoted by $e^*(\delta, m)$. With the remaining probability, the opportunistic type non-strategically exerts no effort. Having these in hand, the equilibrium probability of the opportunistic type's performance being high given $m = \omega$ is

$$\mathbb{E}[y_o^* \mid m = \omega] := a \int_0^\Delta e^*(\delta, \omega) \frac{m^*(\delta, \omega)}{\Delta \mathbb{E}[m^*(\delta, \omega)]} d\delta.$$

First, suppose that voter i thinks that the incumbent supports the truth. 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 $\tilde{y}_{1i}$:

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

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. Hence, with probability $ae_1 + q_i(1 - ae_1)$, $\tilde{y}_{1i} = 1$. Therefore, we have

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

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, regardless of $\tilde{y}_{1i}$,

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

Therefore,

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

Using these values, we obtain $\mathbb{E}[p_i \mid 1, e_1]$ and $\mathbb{E}[p_i \mid 0, e_1]$. Furthermore, by substituting them into the reelection probability, we obtain the lemma. □

Proof of Lemma 3 Letting f be a density function,

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

Note that even if F is discrete, the same formula is derived. In addition,

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

By substituting them, we have

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

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

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

Preliminary for the Proofs of Propositions 1 and 2 We prove the following lemma:

Lemma 4. *For any σ^2, 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 effort level given $m = 1$. Because it is the solution to

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

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

Note that this lies in $(0, 1/a)$ from Assumption 1. In our model, the domain of δ is $(0, \Delta)$, but we also define $e^*(0, 1)$ as that given by the above formula.

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

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

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

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

By taking the first-order condition, we obtain

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

Note that this lies in $(0, 1/a)$ from Assumption 1. In our model, the domain of δ is $(0, \Delta)$, but we also define $e^*(0, 0)$ as that given by the above formula. Substituting this into $U(0, e)$ yields the incumbent's value function when $m = 0$:

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

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

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

which holds from Assumption 5 and Lemma 3. Therefore, we obtain the desired result. □

Proof of Proposition 1 It suffices to prove that $\bar{\delta} = 0$. For this purpose, it suffices to prove that $U(1, e^*(\delta, 1) \mid \delta) \geq U(0, e^*(\delta, 0) \mid \delta)$ when $\delta = 0$. When $\delta = 0$,

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

where the third line is obtained by substituting the values of $\mathbb{E}[q_i \mid I]$ and $\mathbb{E}[q_i \mid M]$, and $\mathbb{E}[q^2] = \mathbb{E}[q]^2 + \sigma^2$. Therefore, $U(1, e^*(0, 1) \mid 0) \geq U(0, e^*(0, 0) \mid 0)$, implying that $\bar{\delta} = 0$. □

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 δ 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. We show that there is at most one δ such that $V(\delta) = 0$. To this end, it suffices to prove that $V'(\delta) > 0$ when $V(\delta) \geq 0$. Letting $P^*(\delta, m \mid \bar{\delta})$ be $P^*(m, e^*(\delta, m) \mid \bar{\delta})$, we have

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

where the second equality comes from $\frac{\partial U}{\partial e}(1, e^*(\delta, 1 \mid \delta) \mid \delta) = \frac{\partial U}{\partial e}(0, e^*(\delta, 0 \mid \delta) \mid \delta) = 0$ (i.e., the envelope theorem).

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

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

$$\begin{aligned}
&\frac{\partial P^*}{\partial \delta}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \delta}(\delta, 0 \mid \delta) \\
&= \psi \left\{ [\kappa + (1-\kappa)(1-\mathbb{E}[q])] \left[ae^*(\delta, 1 \mid \delta) + \mathbb{E}[q_i \mid I](1 - ae^*(\delta, 1 \mid \delta))\right] \right. \\
&\quad \left. -(1-\kappa)\mathbb{E}[q] \left[ae^*(\delta, 0 \mid \delta) + \mathbb{E}[q_i \mid M](1 - ae^*(\delta, 0 \mid \delta))\right] \right\} \\
&\quad \times \frac{\partial}{\partial \delta} \left(\frac{1}{1 + (1-\varepsilon)a \int_{\delta}^{\Delta} \frac{e^*(1,\tilde{\delta})}{\Delta} d\tilde{\delta}} \right).
\end{aligned} \tag{7}$$

Here, the last term is positive. In addition, by regarding the first term, we have

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

so that the first term of (7) is non-negative. Taken together, we have (7) ≥ 0.

(ii). $e^*(\delta, 1 \mid \delta) > e^*(\delta, 0 \mid \delta)$. From Lemma 4, it is shown that

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

This condition holds because $\kappa + (1-\kappa)(1-\mathbb{E}[q]) > (1-\kappa)\mathbb{E}[q]$ from Assumption 5 and $\mathbb{E}[q_i \mid I] \leq \mathbb{E}[q_i \mid M]$ from Lemma 3. 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 δ 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 δ 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. □

Proof of Proposition 2 (ii) We prove this step by step.

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 $V(\delta) \geq 0$.

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

Hence, it suffices to prove $\frac{\partial P^*}{\partial \sigma^2}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \sigma^2}(\delta, 0 \mid \delta) < 0$. A simple calculation yields:

$$\begin{aligned}&\frac{\partial P^*}{\partial \sigma^2}(\delta, 1 \mid \delta) - \frac{\partial P^*}{\partial \sigma^2}(\delta, 0 \mid \delta) \\ &\qquad = \psi \left\{[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\left[ae^*(\delta, 1 \mid \delta) + \mathbb{E}[q_i \mid I](1 - ae^*(\delta, 1 \mid \delta))\right]\right. \\ &\qquad\quad \left. -(1-\kappa)\mathbb{E}[q]\left[ae^*(\delta, 0 \mid \delta) + \mathbb{E}[q_i \mid M](1 - ae^*(\delta, 0 \mid \delta))\right]\right\} \\ &\qquad\quad \times \frac{\partial}{\partial \mathbb{E}[y_o^* \mid m = \omega]}\left(\frac{\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}\right)\frac{\partial \mathbb{E}[y_o^* \mid m = \omega]}{\partial \sigma^2} \\ &\qquad\quad + (1 - ae^*(\delta, 1 \mid \delta))\frac{\partial \mathbb{E}[q_i \mid I]}{\partial \sigma^2} - (1 - ae^*(\delta, 0 \mid \delta))\frac{\partial \mathbb{E}[q_i \mid M]}{\partial \sigma^2}.\end{aligned} \tag{8}$$

Here, the last two terms are obviously negative because $\frac{\partial \mathbb{E}[q_i|I]}{\partial \sigma^2} < 0$ and $\frac{\partial \mathbb{E}[q_i|M]}{\partial \sigma^2} > 0$ holds. Furthermore, the first bracket is non-negative because $V(\delta) \geq 0$ is assumed. Hence, (8) 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 by contradiction. Suppose that $\frac{\partial \mathbb{E}[y_o^*|m=\omega]}{\partial \sigma^2} < 0$ holds for some σ^2 so that $\mathbb{E}[y_o^* \mid m = \omega]$ is decreasing in σ^2 for some $\sigma^2 \in [b_1, b_2]$ where $0 < b_1 < b_2$. From the derivation of $e^*(\delta, 1)$ in Lemma 4, this implies that $e^*(\delta, 1)$ is increasing in $\sigma^2 \in$

$[b_1, b_2]$ for any δ.[25] However, this further implies that $\mathbb{E}[y_o^* \mid m = \omega]$ is decreasing in $\sigma^2 \in [b_1, b_2]$, which is a contradiction. Therefore, $\frac{\partial \mathbb{E}[y_o^* | m=\omega]}{\partial \sigma^2} \geq 0$, implying that (8) is negative.

Step 2. Now, we are ready to prove the proposition. Let $\bar{\delta}$ given σ^2 be $\bar{\delta}(\sigma^2)$ and $V(\delta)$ given σ^2 be $V(\delta \mid \sigma^2)$. Prove by contradiction. Suppose that $\bar{\delta}$ is non-increasing in $\sigma^2 \in [b_1, b_2]$ where $0 < b_1 < b_2$. $V(\delta \mid \sigma^2) \geq 0$ holds for any $\delta(\geq \bar{\delta}(b_1))$ and $\sigma^2 \in [b_1, b_2]$. Hence, from Step 1, $0 = V(\bar{\delta}(b_1) \mid b_1) > V(\bar{\delta}(b_1) \mid b_2)$. This implies that $\bar{\delta}(b_2) > \bar{\delta}(b_1)$, which is a contradiction. Therefore, we obtain the desired result. □

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 4. 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 to that in the proof of Lemma 4 yields $e^*(\delta, 1) = \delta a b \psi R_1$. Hence, it suffices to derive the value of R_1. Now,

$$\begin{aligned} R_1 =& \mathbb{E}[p_i \mid m, y_1 = 1] - \mathbb{E}[p_i \mid m, y_1 = 0] \\ =& \frac{[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = 1]} - \frac{[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\mathbb{E}[q_i \mid I]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = 1]}. \end{aligned} \tag{9}$$

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

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

and $\bar{p} = \frac{1}{2-\varepsilon}$. By substituting them into (9), we obtain the following equation:

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

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

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

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

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

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

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

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

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

This is increasing in δ. □

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, (9) does not change. Only $\mathbb{E}[y_o^* \mid m = \omega]$ and $\bar{p}$ change:

$$\mathbb{E}[y_o^* \mid m = \omega] = a\int_{\bar{\delta}}^{\Delta} e^*(\delta, 1)\frac{1}{\Delta - \bar{\delta}}d\delta = a\int_0^{\Delta}\frac{\delta ab\psi R_1}{\Delta - \bar{\delta}}d\delta = \frac{a^2 b\psi(\Delta + \bar{\delta})}{2}R_1; \quad (10)$$

$$\bar{p} = \frac{\Delta}{\Delta + (1-\varepsilon)(\Delta - \bar{\delta})}. \quad (11)$$

Step 1. We first derive $e^*(\delta, 1)$. By substituting (10) and (11) into equation (9), we have

$$(1-\varepsilon)\frac{a^2 b\psi(\Delta^2 - \bar{\delta}^2)}{2}R_1^2 + \Delta R_1 - [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i \mid I])\Delta = 0. \quad (12)$$

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

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

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

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

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

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

Furthermore, this is obviously increasing in δ.

Step 2. Next, we derive the value of $e^*(\delta, 0)$. Given that $m = 0$, $e^*(\delta, 0) = \delta ab\psi R_0$. Hence, it suffices to derive the value of R_0. Now,

$$R_0 = \frac{(1-\kappa)\mathbb{E}[q](1-\mathbb{E}[q_i \mid M])\Delta}{\Delta + 0.5(1-\varepsilon)(\Delta^2 - \bar{\delta}^2)a^2 b\psi R_1}. \tag{13}$$

By substituting the value of R_1 into this, we obtain

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

This is increasing in δ.

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

Proof of Proposition 4 (i) is straightforward because e^{**} is decreasing in $\mathbb{E}[q^\tau \mid I]$ and higher σ^2 reduces $\mathbb{E}[q^\tau \mid I]$ from Lemma 3. Thus, we focus on (ii).

Step 1. From Proposition 3, $\lim_{\delta \nearrow \bar{\delta}} e^{**}(\delta) < \lim_{\delta \searrow \bar{\delta}} e^{**}(\delta)$, and $\bar{\delta}(\nu + \varsigma) > \bar{\delta}(\nu)$ from Proposition 2. Hence, $e^{**}(\delta \mid \nu + \varsigma) < e^{**}(\delta \mid \nu)$ for $\delta \in [\bar{\delta}(\nu), \bar{\delta}(\nu + \varsigma))$.

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

$$\frac{dR_1(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} = \frac{\partial R_1}{\partial \bar{\delta}}(\sigma^2, \bar{\delta}(\sigma^2))\frac{\partial \bar{\delta}}{\partial \sigma^2} + \frac{\partial R_1}{\partial \sigma^2}(\sigma^2, \bar{\delta}(\sigma^2)) > 0. \tag{14}$$

If this holds, $e^*(\delta, 1)$ is increasing in σ^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 (12), 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 σ^2 for equation (12), we obtain the following:

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

Combining them yields (14), which completes the proof of Step 2.

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

$$\frac{dR_0(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} = \frac{\partial R_0}{\partial R_1}\frac{dR_1(\sigma^2, \bar{\delta}(\sigma^2))}{d\sigma^2} + \frac{\partial R_0}{\partial \sigma^2} < 0. \tag{15}$$

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

First, from equation (13), 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 $\mathbb{E}[q_i \mid M]$ and $\mathbb{E}[q_i \mid M]$ is increasing in σ^2. Hence, $\frac{\partial R_0}{\partial \sigma^2} < 0$.

Combining them yields (15), which completes the proof of Step 3.

From Steps 1-3, we obtain the desired result. □

B For Online Publication: Online Appendix for “Misinformation in Representative Democracy: The Role of Heterogeneous Confirmation Bias"

This online appendix provides detailed discussions on implications, extensions, and modeling assumptions of our baseline model, which were briefly explained in Section 6. Specifically, we discuss (i) the effect of denying the truth on the reelection probability, (ii) the challenger’s message, (iii) the correlation of confirmation bias across issues, (iv) repeated elections with term limits, (v) strategic incentives of the ethical type, (vi) an alternative setting on the misperception probability, (vii) the incumbent’s prior reputation, (viii) a weaker version of confirmation bias, and (ix) an alternative model with policy choice rather than effort choice.

Contents

B.1 Effect on Reelection Probability

In discussing the empirical implications of our model, we have argued that the causal effect of denying the truth given δ is heterogeneous depending on the value of the competence level, δ. In this subsection, we provide the formal proof for this statement.

Proposition B.1. *Suppose that $\sigma^2 > \bar{\sigma}^2$. There exists $\underline{\delta} \in (0, \bar{\delta})$ such that is*

$$P^*(0, e^*(\delta, 0)) - P^*(1, e^*(\delta, 1)) \gtreqless 0 \Leftrightarrow \delta \lesseqgtr \underline{\delta}.$$

Proof. Step 1. First, we prove that $P^*(0, e^*(\delta, 0)) - P^*(1, e^*(\delta, 1))$ is decreasing in δ. From Lemma 2,

$$\begin{aligned}\frac{\partial[P^*(0, e^*(\delta, 0)) - P^*(1, e^*(\delta, 1))]}{\partial \delta} \propto &(1-\kappa)\mathbb{E}[q](1-\mathbb{E}[q_i \mid M])\frac{\partial e^*(\delta, 0)}{\partial \delta} \\ &- [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i \mid I])\frac{\partial e^*(\delta, 1)}{\partial \delta}.\end{aligned} \tag{B.1}$$

Here, combining Assumption 5 with Lemma 3, we have

$$(1-\kappa)\mathbb{E}[q](1-\mathbb{E}[q_i \mid M]) < [\kappa + (1-\kappa)(1-\mathbb{E}[q])](1-\mathbb{E}[q_i \mid I]). \tag{B.2}$$

Furthermore, a simple calculation yields[B1]

$$\frac{\partial e^*(\delta, 0)}{\partial \delta} < \frac{\partial e^*(\delta, 1)}{\partial \delta}. \tag{B.3}$$

By substituting (B.2) and (B.3) into (B.1), it is shown that $P^*(0, e^*(\delta, 0)) - P^*(1, e^*(\delta, 1))$ is decreasing in δ.

[B1] From Proposition 3, $\frac{\partial e^*(\delta,0)}{\partial \delta} < \frac{\partial e^*(\delta,1)}{\partial \delta}$ holds if and only if

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

Because this is equivalent to $e^*(\delta, 1) > e^*(\delta, 0)$, it is always satisfied (see the proof of Proposition 3).

Step 2. Next, we show that for sufficiently small δ, $P^*(0, e^*(\delta, 0)) - P^*(1, e^*(\delta, 1)) > 0$. This is satisfied because when $\delta = 0$,

$$P^*(0, e^*(0, 0)) - P^*(1, e^*(0, 1)) > 0 \Leftrightarrow U(1, e^*(1, 0) \mid 0) - U(0, e^*(0, 0) \mid 0) < 0 \Leftrightarrow \sigma^2 > \bar{\sigma}^2.$$

Step 3. Lastly, we show that $P^*(0, e^*(\bar{\delta}, 0)) - P^*(1, e^*(\bar{\delta}, 1)) < 0$. By definition, for $\delta = \bar{\delta}$,

$$U(1, e^*(\bar{\delta}, 1) \mid \bar{\delta}) = P^*(1, e^*(\bar{\delta}, 1))b - \frac{e^*(\bar{\delta}, 1)^2}{2\bar{\delta}};$$

$$U(0, e^*(\bar{\delta}, 0) \mid \bar{\delta}) = P^*(0, e^*(\bar{\delta}, 0))b - \frac{e^*(\bar{\delta}, 0)^2}{2\bar{\delta}}.$$

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

$$P^*(1, e^*(\bar{\delta}, 1)) - P^*(0, e^*(\bar{\delta}, 0)) = \frac{e^*(\bar{\delta}, 1)^2 - e^*(\bar{\delta}, 0)^2}{2\bar{\delta}b}.$$

From the proof of Proposition 3, $e^*(\delta, 1) > e^*(\delta, 0)$, implying that the right-hand side of the above equality is positive. Hence, $P^*(1, e^*(\bar{\delta}, 1)) - P^*(0, e^*(\bar{\delta}, 0)) > 0$.

Combining Steps 1-3 with the intermediate value theorem leads to the desired result. □

B.2 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). The other settings remain the same. Under this setting, we obtain the following result:

Proposition B.2. *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 δ).*

Proof. 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 $(m_I, \tilde{y}_{1i})$, 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 the majority of votes given (m_I, m_C, e_1) as:

$$\frac{1}{2} + \psi \left(\mathbb{E}[p_i^I \mid m_I, e_1] - \mathbb{E}[p_i^C \mid m_C] \right). \tag{B.4}$$

(i). The incumbent's objective is to maximize (B.4). 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 (B.4), 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 5, $\mathbb{E}[p_i^C \mid m_C = 1] > \mathbb{E}[p_i^C \mid m_C = 0]$, implying that the challenger never denies the truth strategically.

□

Therefore, the challenger has no incentive to deny the truth. The mechanism is as follows. As we have shown in the main analysis, the incumbent denies the truth in order to weaken electoral punishment for low performance. On the contrary, the challenger exerts no effort in the provision of public goods; thus 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 the size of her political support base. Therefore, denying the truth supported by the majority of voters is not optimal for the challenger.

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 this occurrence is unlikely for the incumbent because such politicians lose an election and cannot be the incumbent. In other words, the value of ε, the fraction of the non-strategic opportunistic type, is expected to be much larger for the challenger than for the incumbent. As the probability of denying the truth is given by $\varepsilon + (1-\varepsilon)\mathbb{E}[1 - m^*(\delta, 1)]$, the challenger is more likely to deny the truth when this selection effect dominates the incentive effect presented in the above proposition. Distinguishing the dominant effect will be left for future empirical research.

B.3 Correlation of Confirmation Bias across Issues

The degree of confirmation bias when learning ω and that when learning the incumbent's type may differ even within individuals because the subject of learning differs. To consider this possibility, let us suppose that the degree of confirmation bias when learning ω is q_i^ω and that when learning the incumbent's type is q_i^τ. In particular, we suppose that $q_i^\omega = q_i^\tau$ are the same for all i with probability λ, but q_i^ω and q_i^τ 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.

In this setting, the same results hold except that the value of $\bar{\sigma}^2$ is modified as follows:

Proposition B.3.

$$\bar{\sigma}^2 = \frac{\mathbb{E}[q]\,[1 - 2(1-\kappa)\mathbb{E}[q]]}{2\lambda(1-\kappa)}.$$

Proof. First of all, it can be easily observed that

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

where

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

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

where

$$\mathbb{E}[q_i^\omega \mid M] = \frac{\mathbb{E}[q^2]}{\mathbb{E}[q]}.$$

By replacing $\mathbb{E}[q_i \mid I]$ and $\mathbb{E}[q_i \mid M]$ with $\mathbb{E}[q_i^\tau \mid I]$ and $\mathbb{E}[q_i^\tau \mid M]$, we can apply the proofs of the main analysis. In particular, as in the proof of Proposition 1, we have

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

□

Thus, $\bar{\sigma}^2$ is increasing in the degree of correlation across issues. Furthermore, as λ converges to zero, $\bar{\sigma}^2$ goes to infinity, implying that the incumbent has no incentive to deny the truth when the degree of correlation across issues is quite low.

To see the mechanism, suppose an extreme case wherein the correlation is zero (i.e., $\lambda = 0$). Then, voters believing the false argument about ω have a high confirmation bias when learning the truth on this issue, but their confirmation bias when learning the incumbent's type is not necessarily low. Hence, attracting voters believing the false argument about ω does not help weaken electoral punishment. When the correlation is low, the incumbent has no incentive to deny the truth. This result indicates that the effect of the heterogeneity of confirmation bias depends on whether is is across or within individuals. When it is about the correlation across voters, higher heterogeneity induces strategic misinformation, but when it is about the correlation across issues, the opposite is the case.

B.4 Repeated Elections Model with Term-Limits

Thus far, we have analyzed the two-period model. While it has been widely used in the literature (e.g., Besley, 2006), the reality is that the world does not end after two periods. In this section, we will prove that our results still hold even in an infinitely repeated elections model with term-limits (e.g., Banks and Sundaram, 1998; Smart and Sturm, 2013; Kartik and Van Weelden, 2019*b*).

Setting: For this purpose, we extend our two-period model to the following infinite horizon model with discrtete time, idexed by $t = 0, 1, \ldots$. The common discount factor is $\delta \in (0, 1)$. Each stage game is given as follows:

1. Voters observe a public signal, s_t, about ω_t.

2. The incumbent[B2] sends a message, m_t, about ω_t. In addition, the incumbent exerts an

[B2] In period 0, the incumbent is exogenously given.

effort, e_t, which is unobservable to voters.

3. Voters observe y_t (i.e., whether public goods are successfully provided). Then, they vote for the incumbent or a newly born challenger. The candidate obtaining the majority of votes will become elected as the policymaker (i.e., the incumbent) in period $t + 1$.

We assume two-term limits as in the US presidential election. Thus, if the incumbent in period t has been the policymaker in both periods t and $t - 1$, she cannot be running for the election at the end of period t. In this case, only a newly born challenger is running for an election and she becomes the policymaker in period $t + 1$.

Each politician's type is independently drawn as in the two-period model it is invariant across time. Furthermore, a controversial issue differs across time and thus, ω_t is independently drawn across time.

Hereafter, in line with the literature of repeated elections (e.g., Smart and Sturm, 2013; Kartik and Van Weelden, 2019*b*), we characterize Markov perfect equilibria of this game (i.e., equilibrium strategies depend only on payoff relevant information). More formally, the Markov perfect equilibrium is defined as a perfect Bayesian equilibrium where (i) the incumbent's equilibrium choice about m_t depends only on her type and ω_t; (ii) the incumbent's equilibrium choice about e_t depends only on her type, ω_t, and m_t; and (iii) each voter's voting depends only on his belief on the incumbent's type.

Voters' voting decision: Suppose that the incumbent in period t was first elected at the end of period $t - 1$ so that she can run for the election at the end of period t and can be reelected as the policymaker in period $t + 1$. Let voter i's equilibrium continuation payoff from public goods provision when the challenger is elected as the policymaker in period $t + 1$ be V^C. This is independent of t and the past history because we focus on Markov perfect equilibria. Then, the voter i's equilibrium continuation payoff from public goods provision when the incumbent is elected as the policymaker in period $t + 1$ is given by

$$p_{it} + \delta V^C,$$

where p_{it} is voter i's subjective probability of the incumbent being ethical. To see why this holds, remember that the incumbent cannot be reelected in the next period due to two-term limits; period $t + 1$ is her last term. Hence, she exerts effort in period $t + 1$ only when she is the ethical type, implying that voter i's expected payoff of period $t + 1$ is given by p_{it}.

Furthermore, at the end of period $t+1$, a challenger is elected for sure so that the voters' continuation payoff from the next period is given by δV^C.

Taken together, voter i votes for the incumbent at the end of period t if and only if[B3]

$$p_{it} + \delta V^C + \zeta_{it} + \eta_t \geq V^C \;\Leftrightarrow\; p_{it} + \zeta_i + \eta \geq (1-\delta)V^C.$$

Note that η_t and ζ_{it} are independently and identically distributed across time.

By applying the procedure same as in Section 3.3 to this inequality, we obtain the reelection probability of the incumbent as follows:

$$\frac{1}{2} + \psi\left(\mathbb{E}[p_{it} \mid m_t, e_t] - (1-\delta)V^C\right).$$

Equilibrium: This formula for the reelection probability is basically the same as that in the two-period model.[B4] Therefore, the same analysis is applicable, leading to the following proposition.

Proposition B.4. *In the Markov perfect equilibria, the following holds:*

(i). Suppose that the policymaker in period t has been the policymaker since the beginning of period $t-1$. Then, if she is the opportunistic type and strategic, $e_t^* = 0$ and she is indifferent between $m_t = 0$ and $m_t = 1$.

(ii). Suppose that the policymaker in period t was first elected at the end of period $t-1$. Then, if she is the opportunistic type and strategic, e_t^* and m_t^* are given by those in Propositions 1 and 2.

Proof. (i). In this case, the policymaker will be never reelected. Thus, she maximizes

$$b - \frac{e_t^2}{2\delta}.$$

Therefore, $e_t^* = 0$ and she is indifferent between $m_t = 0$ and $m_t = 1$.

(ii). In this case, if she is reelected, she will choose $e_{t+1} = 0$ and get retired at the end of period $t+1$ from (i). Hence, her objective in period t is to maximize

$$b\frac{e_t^2}{2\delta} + \delta\left[\frac{1}{2} + \psi\left(\mathbb{E}[p_{it} \mid m_t, e_t] - (1-\delta)V^C\right)\right] b.$$

[B3] Similarly with our setting, Kartik and Van Weelden (2019*b*) introduce probabilistic voting into a repeated elections model with term-limits.

[B4] The third term of reelection probability differs, but it is independent of the incumbent's choice on m_t and e_t; thus, it does not matter for the incumbent's maximization problem.

The solution to this maximization problem (i.e., the best response of the opportunistic type) is the same as that in the two-period model because the terms depending on m_t and e_t are exactly the same. Therefore, the equilibrium characterization is the same as that in Propositions 1 and 2.

□

B.5 Strategic Incentive of Ethical Type

Thus far, we have assumed that the ethical type non-strategically supports the truth. While we believe that this is reasonable, a natural question would be what happen if we allow the ethical type to strategically behave. In this section, we will show that our results still hold even in such a case.

For this purpose, suppose that the ethical type is strategic with probability $1 - \varepsilon$. Specifically, if she is strategic, the ethical type maximizes voters' welfare:[B5]

$$\mathbb{E}[y_{i1}] + \mathbb{E}[y_{i2}].$$

Notice that voters' beliefs on the value of ω do not respond to the incumbent's message, m. Hence, in maximizing voters' welfare, it suffices to focus on the public goods provision.

Under this setting, we obtain the following result. Let the ethical type's equilibrium strategy of m and e_t be m^*_E and e^*_{tE}. We focus on equilibria where the ethical type's equilibrium strategy is independent of δ because it is payoff-irrelevant for them.

Proposition B.5. *The following holds:*

(i). In every equilibrium, $e^*_{1E} = e^*_{2E} = 1/a$.

(ii). There is an equilibrium where $m^*_E = 1$. In such an equilibrium, the opportunistic type's equilibrium strategy is characterized by Propositions 1 and 2.

(iii). Suppose that $\sigma^2 = 0$. There is no equilibrium where the ethical type has a strict incentive to choose $m = 0$.

Proof. (i). We first prove that $e^*_{1E} = e^*_{2E} = 1/a$. In period 2, the ethical type maximizes $\mathbb{E}[y_{i2}] = ae_2$; thus, $e^*_{2E} = 1/a$ straightforwardly holds.

[B5] For simplicity, we assume that she does not care about random utility terms (ζ_i and η) in evaluating the welfare.

Furthermore, in period 1, the ethical type maximizes

$$\begin{aligned}\mathbb{E}[y_{i1}] + \mathbb{E}[y_{i2}] =& ae_1 + P^*(m, e_1) \times 1 + (1 - P^*(m, e_1)) \times \frac{1}{2} \\ =& ae_1 + \frac{1 + P^*(m, e_1)}{2} \\ =& ae_1 + \frac{1 + ae_1 P^*(m, y_1 = 1) + (1 - ae_1) P^*(m, y_1 = 0)}{2},\end{aligned} \tag{B.5}$$

where $P^*(m, e_1)$ represents the probability of reelection given (m, e_1), and $P^*(m, y_1)$ represents the probability of reelection given (m, y_1). Note that the first line comes from the fact that the maximum effort will be exerted and the public goods will be successfully provided if she is reelected, but otherwise, it will happen only with probability half (the challenger is the ethical type with probability half).

From (B.5),

$$\frac{\partial(\mathbb{E}[y_{i1}] + \mathbb{E}[y_{i2}])}{\partial e_1} = a + \frac{a}{2}[P^*(m, y_1 = 1) - P^*(m, y_1 = 0)] > 0$$

where the last inequality is satisfied because $P^*(m, y_1 = 1) - P^*(m, y_1 = 0) \geq -1$ holds by the definition of the probability. Therefore, $e^*_{1E} = 1/a$.

(ii). Given that $m^*_E = 1$, the opportunistic type's equilibrium strategy is derived as in Propositions 1 and 2. Furthermore, given that $m^*_E = 1$, the strategic ethical type has no incentive to choose $m = 0$ because $P^*(0, e_1 = 1/a) < P^*(1, e_1 = 1/a)$ holds from Assumption 5. Therefore, (ii) is obtained.

(iii). Lastly, we analyze the case where $\sigma^2 = 0$. We first prove that there is no equilibrium where $m^*_E = 0$. We prove this by contradiction. From (i) and (B.5), the ethical type has an incentive to choose $m = 0$ if and only if

$$P^*(0, e_1 = 1/a) > P^*(1, e_1 = 1/a).$$

We consider two cases. Let $p_I(m, \tilde{y})$ be informed voters' subjective belief given $(m, \tilde{y})$. Note that $p_I(m, 0) = 0$ because the ethical type's y is always equal to one from (i).

Case (a). $p_I^{int}(1) \geq 0.5 \geq p_I^{int}(0)$. As a sub-case, first consider the case where

$p_I^{int}(1) > 0.5 > p_I^{int}(0)$. In this case,

$$P^*(1, e_1) = \frac{1-\psi}{2} + \psi[\kappa + (1-\kappa)(1-\mathbb{E}[q])](ae_1 + (1-ae_1)\mathbb{E}[q])p_I(1,1) + \psi(1-\kappa)\mathbb{E}[q]ae_1(1-\mathbb{E}[q])p_I(0,1);$$

$$P^*(0, e_1) = \frac{1-\psi}{2} + \psi[\kappa + (1-\kappa)(1-\mathbb{E}[q])]ae_1(1-\mathbb{E}[q])p_I(0,1) + \psi(1-\kappa)\mathbb{E}[q](ae_1 + (1-ae_1)\mathbb{E}[q])p_I(1,1).$$

Hence, $P^*(1, e_1 = 1/a) > P^*(0, e_1 = 1/a)$ holds because $p_I(1,1) > p_I(0,1)$. This is a contradiction. The same holds even if either $p_I^{int}(1) = 0.5$ or $p_I^{int}(0) = 0.5$ holds.

Case (b). $p_I^{int}(1) \leq 0.5 \leq p_I^{int}(0)$. As a sub-case, first, consider the case where $p_I^{int}(1) < 0.5 < p_I^{int}(0)$. In this case,

$$P^*(1, e_1) = \frac{1-\psi}{2} + \psi[\kappa + (1-\kappa)(1-\mathbb{E}[q])]ae_1(1-\mathbb{E}[q])p_I(1,1) + \psi(1-\kappa)\mathbb{E}[q](ae_1 + (1-ae_1)\mathbb{E}[q])p_I(0,1);$$

$$P^*(0, e_1) = \frac{1-\psi}{2} + \psi[\kappa + (1-\kappa)(1-\mathbb{E}[q])](ae_1 + (1-ae_1)\mathbb{E}[q])p_I(0,1) + \psi(1-\kappa)\mathbb{E}[q]ae_1(1-\mathbb{E}[q])p_I(1,1).$$

Hence, $P^*(1, e_1) < P^*(0, e_1)$ holds for any e_1. Therefore, the strategic opportunistic type chooses $m = 0$ for any δ. However, if so, $p_I^{int}(1) = 1$ because only the non-strategic ethical type chooses $m = 1$. The same holds even if either $p_I^{int}(1) = 0.5$ or $p_I^{int}(0) = 0.5$ holds. This is a contradiction.

From cases (a) and (b), we conclude that $P^*(0, e_1 = 1/a) > P^*(1, e_1 = 1/a)$ never holds, leading to (iii).

□

First, (i) in the above proposition argues that the ethical type exerts the maximum effort even if she is strategic. This is straightforward because the ethical type attempts to maximize the probability of public goods being successfully provided.

(ii) and (iii) are concerning whether to deny the truth. (ii) shows that there is an equilibrium where the ethical type never denies the truth and the opportunistic type's strategy is characterized as in the main analysis in such an equilibrium. To see the mechanism, assume that candidates follow this equilibrium strategy. As our main analysis has shown, given this equilibrium, denying the truth boosts the reelection probability only when the incumbent's level of exerted effort is low. Here, the ethical type exerts the maximum effort (see (i)); thus, denying the truth not boosts but undermines his reelection probability. Hence, the ethical type does not deny the truth. Furthermore, given that the ethical type claims the truth, the situation is the same as in the main analysis; thus, the same strategy of the opportunistic type constitutes an equilibrium.

However, (ii) does not rule out a possibility for the existence of another equilibrium. (iii) focuses on the case without heterogeneity of confirmation bias and shows that there is no equilibrium where the ethical type has a strict incentive to choose $m = 0$. This implies that with an arbitrarily small lying cost, the ethical type has no incentive to deny the truth. Together with (ii), this indicates that only the equilibrium characterized in the main analysis is a reasonable equilibrium when $\sigma^2 = 0$.

The main message of our main analysis was that the incumbent may deny the truth only when the heterogeneity of confirmation bias exists. (ii) and (iii) indicate that this property still holds. On the one hand, when $\sigma^2 = 0$, the incumbent has no strict incentive to strategically deny the truth (see (iii)). On the other hand, when σ^2 is sufficiently large, there is an equilibrium the opportunistic type has the incentive to deny the truth (see (ii)). As such, the main message of our main analysis remains.

B.6 Probability of Misperception

Our setting on confirmation bias is a simplified one where the probability of misperception is constant as long as the prior belief is not perfectly neutral. For instance, when $\alpha_i > 0.5$, voter i misperceives $s_i = 0$ as $s_i = 1$ with probability q_i, where q_i is constant. In reality, however, this is a too simplified assumption. It is more reasonable to assume that the misperception probability depends on the strength of the prior belief.[B6] Formally, suppose that the misperception probability when updating ω is $2r_i|\alpha_i - 0.5|$ and that when updating the incumbent's type given y is $2r_i|p_i^{int} - 0.5|$, where $r_i \in [0, 1]$.[B7] We replace Assumptions 2 and 4 by these probabilities.[B8] In this setting, r_i captures an individual's psychological

[B6] Indeed, Rabin and Schrag (1999) themselves mention that "[w]e assume that the severity of the bias summarized by q does not depend on the strength of the agent's beliefs about which of the two states is more likely. It would be reasonable to expect that q is greater if the agent's beliefs are more extreme" (pp.48-49).

[B7] When the prior is either one or zero, the misperception probability is maximized and takes a value r_i.

[B8] The initial probability of the incumbent's being ethical is half; thus, the misperception probability for the incumbent's message is zero as in Assumption 3.

factor and the probability of misperception is larger as one's belief becomes more extreme. We assume that the variance of r_i is σ^2.

To make notations simplified, in the following, assume that $|\underline{\alpha} - 0.5| = |\bar{\alpha} - 0.5| =: c_\alpha$. In this setting, the size of the political support base when supporting the truth (i.e., the number of informed voters) is given by

$$\kappa + (1 - \kappa)(1 - 2\mathbb{E}[r]c_\alpha),$$

and the expected probability of misperceiving low performance as high among this political support base is

$$2\mathbb{E}[r_i \mid I]|\bar{p} - 0.5|.$$

Similarly, the size of the political support base when denying the truth (i.e., the number of misinformed voters) is given by

$$(1 - \kappa)2\mathbb{E}[r]c_\alpha,$$

and the expected probability of misperceiving low performance as high among this political support base is

$$2\mathbb{E}[r_i \mid M]|\bar{p} - 0.5|.$$

Consequently, the reelection probability when supporting the truth is

$$P^*(1, e) = \frac{1 - \psi}{2} + \psi[\kappa + (1 - \kappa)(1 - 2\mathbb{E}[r]c_\alpha)] \frac{[ae + 2\mathbb{E}[r_i \mid I]|\bar{p} - 0.5|(1 - ae)]\, \bar{p}}{\bar{p} + (1 - \bar{p})\mathbb{E}[y_o^* \mid m = \omega]},$$

whereas that when denying the truth is

$$P^*(0, e) = \frac{1 - \psi}{2} + \psi(1 - \kappa)2\mathbb{E}[r]c_\alpha \frac{[ae + 2\mathbb{E}[r_i \mid M]|\bar{p} - 0.5|(1 - ae)]\, \bar{p}}{\bar{p} + (1 - \bar{p})\mathbb{E}[y_o^* \mid m = \omega]}.$$

In this alternative model, the probability of misperceiving low performance as high depends on $\bar{p}$, which further depends on the incumbent's equilibrium strategy. Hence, the reelection probability is more complicated than in the main analysis. While this is the case, we can, at least, easily derive the condition for whether there is an equilibrium such that the opportunistic-type never strategically denies the truth. As we have shown in Lemma 4 and the proof of Proposition 1, such an equilibrium exists if and only if the least competent incumbent has no incentive to deny the truth; that is,

$$U(1, e^*(0, 1) \mid 0) \geq U(0, e^*(0, 0) \mid 0) \Leftrightarrow P^*(1, 0) \geq P^*(0, 0).$$

By calculating this condition, we obtain the following proposition:

Proposition B.6. *If and only if*

$$\sigma^2 \leq \frac{\mathbb{E}[r](1-4(1-\kappa)c_\alpha \mathbb{E}[r])}{4(1-\kappa)c_\alpha},$$

there exists an equilibrium where the opportunistic-type incumbent always supports the truth ($m^*(\delta, 1) = 1$) as long as she is strategic.

Proof. We first derive the values of $\mathbb{E}[r_i \mid I]$ and $\mathbb{E}[r_i \mid M]$.

$$\begin{aligned}\mathbb{E}[r_i \mid I] =& \frac{\kappa}{\kappa + (1-\kappa)(1-\mathbb{E}[r]2c_\alpha)}\mathbb{E}[r] + \frac{(1-\kappa)(1-\mathbb{E}[r]2c_\alpha)}{\kappa + (1-\kappa)(1-\mathbb{E}[r]2c_\alpha)}\int_0^1 r_i \frac{(1-2r_i c_\alpha)f(r_i)}{1-\mathbb{E}[r]2c_\alpha} dr_i \\ =& \frac{\kappa\mathbb{E}[r] + (1-\kappa)(\mathbb{E}[r] - 2c_\alpha \mathbb{E}[r^2])}{\kappa + (1-\kappa)(1-2c_\alpha\mathbb{E}[r])}. \end{aligned} \tag{B.6}$$

Similarly,

$$\mathbb{E}[r_i \mid M] = \int_0^1 r_i \frac{r_i 2c_\alpha f(r_i)}{\mathbb{E}[r]2c_\alpha} dq_i = \frac{\mathbb{E}[r^2]}{\mathbb{E}[r]}. \tag{B.7}$$

Now,

$$P^*(1,0) \geq P^*(0,0) \Leftrightarrow [\kappa + (1-\kappa)(1-2\mathbb{E}[r]c_\alpha)]\mathbb{E}[r_i \mid I] \geq (1-\kappa)2\mathbb{E}[r]c_\alpha \mathbb{E}[r_i \mid M]. \tag{B.8}$$

By substituting (B.6) and (B.7) into (B.8), we have

$$\begin{aligned}(B.8) \Leftrightarrow & \kappa\mathbb{E}[r] + (1-\kappa)(\mathbb{E}[r] - 2c_\alpha\mathbb{E}[r^2]) \geq (1-\kappa)2c_\alpha\mathbb{E}[r^2] \\ \Leftrightarrow & \sigma^2 \leq \frac{\mathbb{E}[r](1-4(1-\kappa)c_\alpha\mathbb{E}[r])}{4(1-\kappa)c_\alpha}.\end{aligned}$$

Therefore, we obtain the desired result. □

As in the main analysis, no one strategically denies the truth if and only if the degree of heterogeneity falls below a threshold. In this sense, a qualitatively same property holds even in this alternative setting. Furthermore, the threshold converges to $\bar{\sigma}^2$ as $c_\alpha \to 0.5$. Hence, even qualitatively, the threshold value takes a similar value when the prior belief on the state of the world is extreme.

B.7 Incumbent's Prior Reputation

Thus far, we have assumed that the prior probability of the incumbent being ethical is half. In this subsection, we discuss how results would change if we omit this assumption. Specifically, we focus on the case where the prior probability of the incumbent being ethical type, denoted by p, is higher than half. The case where the prior probability is less than half can be analyzed in a similar way.

In our main analysis, voters update their belief on the incumbent's type correctly in response to the incumbent's message because they have a neutral prior (see Assumption 3). However, this no longer holds in the present case. To analyze this issue, we adopt the alternative setting on confirmation bias that was introduced in Online Appendix B.6. Specifically, we assume that the misperception probability given m is given by $2r_i|p - 0.5|$. The advantage of this alternative model is that the effect of the incumbent's prior reputation is continuous at 0.5. We denote $|p - 0.5|$ by c_p.

Reelection probability when supporting the truth: We first consider the case where the incumbent supports the truth. In this case, the incumbent's message is consistent with the prior for informed voters; thus, they correctly perceive the incumbent's message. As a result, $p_I^{int}(1) = \bar{p}$. On the other hand, misinformed voters regard the incumbent's message as denying the truth, which contradicts with their prior on the incumbent's type. Hence, misinformed voter i misperceives the incumbent's message as claiming the truth with probability $2r_i c_p$. As a result, for misinformed voter i, $p_M^{int}(1) = 0$ with probability $1 - 2r_i c_p$, whereas it is $\bar{p}$ with the remaining probability. Taken together, the incumbent's political support base consists of both informed voters and a part of misinformed voters. Specifically, the size of the political support base is given by

$$S(1) := \kappa + (1 - \kappa)(1 - 2\mathbb{E}[r]c_\alpha) + (1 - \kappa)2\mathbb{E}[r]c_\alpha 2\mathbb{E}[r_i \mid M]c_p,$$

and the average confirmation bias among this support base is given by

$$A(1) := \frac{[\kappa + (1 - \kappa)(1 - 2\mathbb{E}[r]c_\alpha)]\mathbb{E}[r \mid I] + (1 - \kappa)4\mathbb{E}[r]c_\alpha\mathbb{E}[r_i \mid M]^2 c_p}{\kappa + (1 - \kappa)(1 - 2\mathbb{E}[r]c_\alpha) + (1 - \kappa)4\mathbb{E}[r]c_\alpha\mathbb{E}[r_i \mid M]c_p}.$$

By replacing the political support base and the average confirmation bias among the support base in $P^*(1,e)$ with these values, we obtain

$$P^*(1,e) = \frac{1-\psi}{2} + \psi S(1) \frac{\left[ae + 2A(1)c_p(1-ae)\right]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}.$$

Reelection probability when denying the truth: We next consider the case where the incumbent denies the truth. In this case, misinformed voters regard this message as supporting the truth, which is consistent with their prior on the incumbent's type: thus, they perceive the incumbent's message as supporting the truth with probability one. As a result, $p_M^{int}(1) = \bar{p}$. On the other hand, for informed voters, the incumbent's message contradicts with their prior on the incumbent's type. Hence, informed voter i misperceives the incumbent's message as claiming the truth with probability $2r_i c_p$. As a result, for misinformed voter i, $p_I^{int}(1) = 0$ with probability $1 - 2r_i c_p$, whereas it is $\bar{p}$ with the remaining probability. Taken together, the incumbent's political support base consists of both misinformed voters and a part of informed voters. Specifically, the size of the political support base is given by

$$S(0) := [\kappa + (1-\kappa)(1 - 2\mathbb{E}[r]c_\alpha)]2\mathbb{E}[r_i \mid I]c_p + (1-\kappa)2\mathbb{E}[r]c_\alpha,$$

and the average confirmation bias among this support base is given by

$$A(0) := \frac{[\kappa + (1-\kappa)(1 - 2\mathbb{E}[r]c_\alpha)]2\mathbb{E}[r_i \mid I]^2 c_p + (1-\kappa)2\mathbb{E}[r]\mathbb{E}[r_i \mid M]}{[\kappa + (1-\kappa)(1 - 2\mathbb{E}[r]c_\alpha)]2\mathbb{E}[r_i \mid I]c_p + (1-\kappa)2\mathbb{E}[r]c_\alpha}.$$

By replacing the political support base and the average confirmation bias among the support base in $P^*(0,e)$ with these values, we obtain

$$P^*(0,e) = \frac{1-\psi}{2} + \psi S(0) \frac{\left[ae + 2A(0)c_p(1-ae)\right]\bar{p}}{\bar{p} + (1-\bar{p})\mathbb{E}[y_o^* \mid m = \omega]}.$$

Whether to deny the truth: Given this derivation, we are now ready to analyze whether the opportunistic type strategically denies the truth. As in the Online Appendix B.6, there is an equilibrium where no incumbent strategically denies the truth if and only if

$$U(1, e^*(0,1) \mid 0) \geq U(0, e^*(0,0) \mid 0) \Leftrightarrow P^*(1,0) \geq P^*(0,0) \Leftrightarrow S(1)A(1) \geq S(0)A(0).$$

This condition is much more complicated than the original one, but it can be easily observed that a similar property still holds under a certain condition. To argue this point, let us emphasize that, as c_p goes to zero and c_α goes to half, $S(m)$ converges to the size of the political support base in the main analysis and $2A(m)c_p$ also converges to the average confirmation bias among the support base in the main analysis. Hence, when c_p is close to zero and c_α is close to 0.5, $S(1)A(1) < S(0)A(0)$ holds for $\sigma^2 > \bar{\sigma}^2$. That is, when c_p is close to zero and c_α is close to 0.5, the equilibrium without strategic misinformation does not exist in the presence of large heterogeneity of confirmation bias. Overall, this result indicates that as long as the initial reputation of the incumbent is close to neutral and the initial belief on the state of the world is extreme, the same result holds under the modified setting of confirmation bias.

Another setting and further discussion: Having said that, we want to emphasize that our results could be applicable to any initial reputation if we change the interpretation on updating due to confirmation bias.

Empirical studies have shown that people are captured by confirmation bias when the received information has room for multiple interpretations (Lord, Ross and Lepper, 1979; Fryer, Harms and Jackson, 2019, e.g.,). For example, when medical experts release a statement that taking the vaccine is safe, people with anti-vaccine attitudes can believe in conspiracy theories and come to think that the truth is the opposite. The same is true for the incumbent's performance evaluation on the provision of public goods such as economic performance. Whether the economy becomes improved can be measured by multiple factors including economic growth, unemployment rates, and inflation rates. Even if the situation suggests that the economy becomes worse in total, voters with pro-incumbent attitudes may come to think that the economy is good by ignoring bad measurements and focusing on good measurements. Therefore, it is quite plausible to assume that voters are captured by confirmation bias when receiving s or y.

On the contrary, the same might not be true for the case when observing the incumbent's message. To see this, suppose that the incumbent claims that taking the vaccine is unsafe. This is a simple message; thus it would be difficult to interpret that the incumbent claims the opposite argument. Therefore, as for learning about the incumbent's type, it might not be so reasonable to assume that voters misinterpret the incumbent's message to confirm the prior belief. If this is the case, for any p, we should assume that voters never misinterpret the incumbent's message when learning the incumbent's type.

Under this alternative setting, voters update their belief on the incumbent's type correctly in response to the incumbent's message. Therefore, the arguments in our main analysis straightforwardly hold. In this sense, our results could be applicable to any initial reputation if we change the setting on confirmation bias.

B.8 Weaker Version of Confirmation Bias

Thus far, we have assumed that voters misperceive the belief-inconsonant information as the belief-consonant information. Another possibility is that voters simply ignore the belief-inconsonant information rather than misinterpreting it as the opposite information. In this subsection, we examine whether the results hold in this alternative setting on confirmation bias.

For this purpose, we assume the following biased updating. When voter i receives a belief-inconsonant signal, he ignores the signal and does not update the prior belief with probability q_i. With the remaining probability, he literally perceives the signal and updates the belief based on the Bayes rule. The other settings remain the same.

Informed and misinformed voters: At the beginning of the game, voters observe the public signal on ω, which conveys the truth (i.e., $s = 1$).

On the one hand, those with $\alpha_i = \bar{\alpha}$ literally perceive the signal and learn that $\omega = 1$. On the other hand, those with $\alpha_i = \underline{\alpha}$ literally perceive the signal and learn that $\omega = 1$ only with probability $1 - q_i$. Hence, in total, $\kappa + (1 - \kappa)(1 - \mathbb{E}[q])$ fraction of voters are informed as in the main analysis.

With probability q_i, those with $\alpha_i = \underline{\alpha}$ ignore the signal and do not update the prior. Hence, they believe that $\omega = 1$ with probability $\underline{\alpha}$. These are misinformed voters, but the difference from the main analysis is that their subjective belief on the probability of ω being one is not zero but $\underline{\alpha}$ in the current model.

Incumbent's message and reputation: Suppose that the incumbent claims the truth (i.e., $m = 1$). In this case, informed voters update the incumbent's reputation as

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

whereas misinformed voters update it as

$$p_M^{int}(1) = \underline{p}_M := \frac{\underline{\alpha}}{\underline{\alpha}\{1 + (1 - \varepsilon)\mathbb{E}[m^*(\delta, 1)]\} + (1 - \underline{\alpha})\{\varepsilon + (1 - \varepsilon)\mathbb{E}[m^*(\delta, 0)]\}} < 0.5.$$

Next, suppose that the incumbent claims the truth (i.e., $m = 0$). In this case, informed voters update the incumbent's reputation as

$$p_I^{int}(0) = 0,$$

whereas misinformed voters update it as

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

Reelection probability: Based on the above discussion, the incumbent's expected average reputation given (m, e) is obtained as follows:

$$\mathbb{E}[p_i \mid 1, e_1] = [\kappa + (1-\kappa)(1 - \mathbb{E}[q])] \left\{ ae_1 \frac{\bar{p}_I}{\bar{p}_I + (1-\bar{p}_I)\mathbb{E}[y_o^* \mid m = \omega]} + \mathbb{E}[q_i \mid I](1 - ae_1)\bar{p}_I \right\}$$
$$+ (1-\kappa)\mathbb{E}[q]ae_1 \left\{ (1 - \mathbb{E}[q \mid M]) \frac{\underline{p}_M}{\underline{p}_M + (1-\underline{p}_M)\mathbb{E}[y_o^* \mid m = \omega]} + \mathbb{E}[q \mid M]\underline{p}_M \right\};$$

$$\mathbb{E}[p_i \mid 0, e_1] = (1-\kappa)\mathbb{E}[q] \left\{ ae_1 \frac{\bar{p}_M}{\bar{p}_M + (1-\bar{p}_M)\mathbb{E}[y_o^* \mid m = \omega]} + \mathbb{E}[q_i \mid I](1 - ae_1)\bar{p}_M \right\}.$$

By substituting them into (6), we obtain $P^*(m, e)$.

Equilibrium: Given this derivation, we are now ready to analyze whether the opportunistic type strategically denies the truth. As in the Online Appendix B.6, there is an equilibrium where no incumbent strategically denies the truth if and only if

$$U(1, e^*(0,1) \mid 0) \geq U(0, e^*(0,0) \mid 0) \Leftrightarrow P^*(1,0) \geq P^*(0,0) \Leftrightarrow \mathbb{E}[p_i \mid 1, 0] \geq \mathbb{E}[p_i \mid 0, 0].$$

This condition is complicated in general, but it is much simplified when $\underline{\alpha} \to 0$. In such a case, $\bar{p}_M \to \bar{p}_I$ and $\underline{p}_M \to 0$. Therefore,

$$\mathbb{E}[p_i \mid 1, 0] \geq \mathbb{E}[p_i \mid 0, 0] \Leftrightarrow [\kappa + (1-\kappa)(1 - \mathbb{E}[q])]\mathbb{E}[q_i \mid I] \geq [(1-\kappa)\mathbb{E}[q]]\mathbb{E}[q_i \mid M]$$
$$\Leftrightarrow \sigma^2 \leq \bar{\sigma}^2.$$

which is the condition same as in the main analysis. In summary, when $\underline{\alpha} \to 0$, the condition is reduced to that obtained in the main analysis. Without the assumption that $\underline{\alpha} \to 0$, the condition becomes much more complicated, but a qualitatively similar result would hold.

In this sense, our main results do not hinge on the assumption that voters misinterpret a belief-inconsonant signal as the opposite signal.

Proposition B.7. *Suppose that $\underline{\alpha} \to 0$. If and only if $\sigma^2 \leq \bar{\sigma}^2$, there exists an equilibrium where the opportunistic-type incumbent always supports the truth ($m^*(\delta, 1) = 1$) as long as she is strategic.*

B.9 Policy Choice

In the model, we have assumed that the incumbent politician determines the effort level for the provision of public goods. Another conflict of interest between the incumbent politician and voters is concerning policy choice rather than effort choice. Indeed, the literature on electoral accountability has emphasized both types of conflict of interest. Therefore, it would be worthwhile to see whether our results can be extended to the conflict of interest stemming from a biased policy preference. In this subsection, we show that the same properties basically hold even if we consider policy choice as the source of the conflict of interest.

Setting: For this purpose, suppose that the elected politician chooses a policy $x_t \in \{0, 1\}$ in period t. The voter-optimal policy depends on the state of the world, $\omega_t \in \{0, 1\}$; that is, voters' payoff from policymaking in period t is given by $-|x_t - \omega_t|$. ω_t is independently drawn across periods, and the probability of ω_t being one is half. While voters do not know the value of ω_t, it is observable to politicians. Note that the incumbent in period 1 observes ω_1 after sending message m. Before the election, voters observe the value of $y_1 := 1 - |x_1 - \omega_1|$.

The ethical type is a sincere politician in that she truthfully confirms the truth and implements the voter-optimal policy (i.e., $x_t = \omega_t$).

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

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

where $\hat{x}$ represents the politician's ideal policy, which is one with probability half.[B9] Lower δ represents that the politician is more policy-motivated and less office-motivated. Because we do not consider effort choice, δ no longer captures the politician's competence level. We

[B9]The politician's ideal policy is unobservable to voters.

also assume that with probability ε, this type of politician non-strategically denies the truth and chooses her ideal policy $\hat{x}$.

Reelection probability: Let p_i be voter i's subjective probability of the incumbent being ethical at the time of election. In period 2, because there is no reelection concern, the opportunistic type chooses her ideal policy that is equal to the voter-optimal policy only with a probability half. On the contrary, the ethical type chooses the voter-optimal policy. Therefore, voter i votes for the incumbent if and only if

$$p_i + (1 - p_i)\frac{1}{2} + \zeta_i + \eta \geq \frac{1}{2} + \frac{1}{2}\frac{1}{2} \Leftrightarrow \zeta_i \geq \frac{1}{4} - \eta - \frac{1}{2}p_i.$$

By aggregating each voter's decision, we derive the incumbent's winning probability. The probability of the incumbent obtaining the majority of votes given (m, y_1) is

$$P^*(m, y_1) := \Pr\left(\underbrace{\frac{1}{2} + \gamma\left(\frac{1}{2}\mathbb{E}[p_i \mid m, y_1] - \frac{1}{4} + \eta\right)}_{\#\text{ of the incumbent's votes given }\eta} \geq \frac{1}{2}\right) = \frac{1}{2} + \frac{\psi}{2}\left(\mathbb{E}[p_i \mid m, y_1] - \frac{1}{2}\right).$$

Equilibrium: As obtained in the baseline model (Lemma 4), it is shown that there exists $\bar{\delta} \in [0, \Delta]$ such that $m^*(\delta, 1) = 1$ (resp. $m^*(\delta, 1) = 0$) if $\delta > \bar{\delta}$ (resp. $\delta < \bar{\delta}$). That is, the incumbent with lower δ has a larger incentive to deny the truth because the incumbent with lower δ is less likely to implement the voter-optimal policy, inducing a lower performance.

The incumbent with sufficiently small δ always implements her ideal policy. Hence, her expected utility given m is

$$[0.5P^*(m, 1) + 0.5P^*(m, 0)]\, b$$

because $\hat{x}$ matches with the voter-optimal policy only with a probability half. Combining this with the above property leads to the following fact: $\bar{\delta} = 0$ if and only if

$$0.5P^*(1, 1) + 0.5P^*(1, 0) \geq 0.5P^*(0, 1) + 0.5P^*(0, 0).$$

By calculating this condition, we obtain the following proposition:

Proposition B.8. *If and only if*

$$\sigma^2 \leq \frac{1 + (2\kappa - 1)\mathbb{E}[q] - 2(1-\kappa)\mathbb{E}[q]^2}{2(1-\kappa)},$$

there exists an equilibrium where the opportunistic-type incumbent always supports the truth ($m^*(\delta, 1) = 1$) *as long as she is strategic.*

Proof.

$$\begin{aligned}
&0.5P^*(1,1) + 0.5P^*(1,0) \geq 0.5P^*(0,1) + 0.5P^*(0,0) \\
\Leftrightarrow &[\kappa + (1-\kappa)(1-\mathbb{E}[q])]\,(1 + \mathbb{E}[q \mid I]) \geq [(1-\kappa)\mathbb{E}[q]]\,(1 + \mathbb{E}[q \mid M]) \\
\Leftrightarrow &\sigma^2 \leq \frac{1 + (2\kappa - 1)\mathbb{E}[q] - 2(1-\kappa)\mathbb{E}[q]^2}{2(1-\kappa)}.
\end{aligned}$$

□

As in the main analysis, no one strategically denies the truth if and only if the degree of heterogeneity falls below a threshold. In this sense, a qualitatively same property holds even in this alternative setting.