

SIGNALING VICE: WHY LOOKING BAD WINS VOTES^{*}

Egor Bronnikov[†] ® Alexei Zakharov[‡]

[[Check the latest version of the paper.](#)]

June, 2026

Abstract

We use a formal model to analyze why politicians engage in *vice signaling* — a deliberate display of bad behavior, transgressive views, or lack of moral integrity. We show that it is possible even when voters prefer candidates with moral character, as long as they are also wary of the candidates who they believe are inauthentic or capable of concealing their immorality. Vice signaling allows an immoral candidate to present oneself as a “bad but authentic” type, implying that the competition is both immoral and inauthentic. This tactic is especially effective with politically cynical voters who are prone to drawing sweeping conclusions about the self-serving nature of political actors. Raising the reputational cost of signaling vice has an ambiguous effect — improving voter welfare, while leading to more low-integrity candidates getting elected.

Keywords: Authenticity, bad candidates, cynicism, electoral competition, vice signaling.

Word count: ~5500.

^{*}The ® symbol indicates that the authors are in random order and made equal contribution. We are grateful to Alvaro Delgado-Vega, Giorgio Farage, Martin Castillo-Quintana, and Shigeo Hirano for fruitful discussions and valuable comments.

[†]School of Business and Economics, Maastricht University; Political Science Department, University of Chicago; School of Science, Brīvā Universitāte; egor.bronnikov@maastrichtuniversity.nl.

[‡]University of Chicago; al.v.zakharov@gmail.com.

1 Introduction

Conventional wisdom about electoral competition assumes that office-seeking candidates present themselves however voters find most appealing. Yet modern politics is rife with politicians who routinely violate this expectation, openly advertising qualities that voters dislike, while flaunting dishonesty, courting scandal, and breaching norms of decency as deliberate features of their appeal. Donald Trump in the United States, Italy Silvio Berlusconi in Italy, and Rodrigo Duterte in the Philippines are just a few prominent examples. This presents a double puzzle: If a trait is a liability at the ballot box, why would a rational candidate try to claim it? And why do such candidates succeed, despite the high reputational costs of signaling vice?

In this work we argue that, while voters value moral character, they reward a second quality alongside it: authenticity, understood as a candidate’s inability to lie about who he is, and punish the hypocrisy of those who profess virtue while concealing its absence. A polished profession of virtue is therefore inherently ambiguous: it may issue from a genuinely high-integrity candidate, or from a low-integrity one who, worse still, is a hypocrite adept at hiding one’s moral flaws. Advertising one’s flaws while presenting oneself as “bad but authentic” helps persuade the voters that the candidate may be flawed, but is not a hypocrite.

This way we provide an explanation to a pattern common in populist politics: Politicians strive to present themselves as authentic candidates (Luebke and Engelmann, 2026), devoting substantial effort to appearing incapable of strategically managing their public self. Authenticity is highly sought after by the voters (Enli, 2025), so presenting oneself as bad but authentic serves the double purpose of “unmasking” one’s opponents, emphasizing that their professed integrity is inauthentic — a core part of populist political messaging (Sorensen, 2020; Aiolfi, 2022). Our analysis shows that the latter effect is stronger if the public is more cynical and prone to making sweeping, negative judgments; thus we show that the widespread voter cynicism (Cappella and Jamieson, 1997) is a strategic resource that populist candidates exploit.

We show that vice signaling leads to low-integrity candidates getting elected more often — something that has previously surprised researchers (De Vries and Solaz, 2017; Winters and Weitz-Shapiro, 2013); our work contributes to the growing formal theory literature examining the election of bad or low-valence candidates. We assume a simple case of voters who are homogeneous (with polarized voters, a low-valence but politically aligned candidate can be supported, Strayhorn, 2026), risk-neutral (risk-averse and may prefer a candidate with known, noncongruent preferences, Kartik and Van Weelden, 2019), and do not self-select into running for office (Caselli and Morelli, 2004). Unlike Schnakenberg (2021), Padró i Miquel and Snowberg (2012), Mizuno and Okazawa

(2022), or [Buisseret and Prato \(2016\)](#), we do not assume that candidate quality compromises voter ability to discipline candidates. Our logic also differs from [Gailmard \(2022\)](#) who emphasized the role of special interest politics in the majority not wanting to select competent candidates, and [Acemoglu et al. \(2013\)](#), where candidates implement bad policies as costly signaling of congruence with the voters.

The rest of this paper is structured as follows. In Section 2, we discuss the prevalence of hypocrisy discourse in contemporary politics and review evidence that voters penalize politicians whom they perceive as hypocritical. In Section 3 we focus on vice signaling, arguing that it allows a politician to appear authentic while implicitly casting the opposition as hypocritical. Section 4 contains the model and the results. Section 5 discusses the results and concludes.

2 Hypocrisy and authenticity

Our paper builds on a growing literature that interprets populist politics primarily as an anti-elite style of political communication rather than as a fixed ideological program, centered on the moral confrontation between ordinary citizens and a corrupt or hypocritical establishment, and appealing directly to the voters who distrust established political parties ([Mansbridge and Macedo, 2019](#); [Guriev and Papaioannou, 2022](#)).

A central rhetorical move of populist politicians is to portray the establishment not merely as wrong or self-interested, but as betraying the public whose interests they are supposed to serve ([Rooduijn, 2014](#); [Hunger, 2024](#); [Hawkins, 2009](#)). Elites such as the mainstream parties, being accused of representing alien interests (such as the wealthy, the immigrants, foreign powers or supranational bodies) and undermining democracy, are also accused of doing so in secret, all the while hypocritically pretending to represent the people ([Mudde and Kaltwasser, 2017](#); [Muller, 2016](#)).

Hypocrisy is one of the oldest moral charges in Western thought and in the entire tradition of Anglo-American populism from the Jacksonians onward ([Kazin, 1998](#)). The accusation of opponents as publicly professing virtue and expertise while privately violating norms, enriching themselves, and otherwise exempting themselves from the rules they impose, is a key component of the populist response to COVID policies ([Bobba and Hubé, 2021](#); [Nadesan, 2022](#)) or the climate change discourse ([Gunster et al., 2018](#)), where pro-climate change politicians are regularly accused of having a large carbon footprint. Hypocrisy charges are transparent in debates over immigration, as the liberal pro-immigration elites are frequently accused of being insulated from the labor market costs of immigration ([Mudde, 2019](#)), as well as generally conspiring to betray the national community while claiming universal compassion.

People have a strong aversion to politicians claiming a moral high ground but privately violating their professed norms. Psychological experiments show that hypocrites — those who condemn behaviors they themselves engage in — are judged more harshly than non-hypocritical wrongdoers or “honest hypocrites” who openly admit to the behaviors they condemn. A possible explanation is that the hypocrites are perceived to be more misleading than ordinary liars (Jordan et al., 2017; Jordan and Sommers, 2022) and claiming an undeserved moral benefit (Effron et al., 2018b), with the aversion to hypocrisy being strong in the more individualistic Western cultures (Effron et al., 2018a).

Politicians in particular are held to a stricter standard on moral norms that align with their political agenda, and are punished more harshly for violating them (Bhatti et al., 2013). When the charges of elite rule-breaking turn out to be genuine, the consequences for political and institutional trust can be severe. A stark example is the Dominic Cummings scandal in the UK, when Prime Minister Boris Johnson’s chief adviser and a key architect of the COVID lockdown rules was found to have violated the stay-at-home rules he had helped design (Fancourt et al., 2020). Relatedly, revelations of electoral fraud in an autocratic setting have a negative effect on regime trust, with the stronger effect on political loyalists who initially believed the regime to be fair (Reuter and Szakonyi, 2021).

The hypocrisy penalty implies a corresponding premium: Voters should reward candidates whose self-presentation seems immune to such suspicion. A growing body of research focuses on politician authenticity — a quality that can be thought of as an incapacity to strategically misrepresent one’s preferences — that is perceived to be distinct from honesty or integrity (Bailey and Iyengar, 2022; Huttunen and Lewandowsky, 2026; Hahl et al., 2018). Online survey experiments suggest that authenticity is valued even when controlling for other aspects of candidate image, including competence, integrity, or sincerity (Valgarðsson et al., 2025; Luebke and Steffan, 2025). This separability is critical for our argument: it implies that voters can in principle reward a candidate for being authentic even when the candidate is not honest or virtuous in the conventional sense.

Luebke (2021) identifies several ways politicians use to communicate authenticity: Showing capacity for spontaneous and impulsive action (including being politically incorrect); exposing one’s intimate motives to the public; appearing consistent across space and time; and demonstrating they are ordinary and imperfect. Authenticity messaging is especially important to populist politics — Fieschi (2019) even argued that it is “the idea that holds it all together”. Populist politicians can go to great lengths to demonstrate authenticity, such as by appearing amateurish and unpolished (Enli, 2017) and denying the use of campaign professionals (Moffitt, 2016). All of these tactics ulti-

mately serve to demonstrate the politician’s incapacity to hide one’s true self from the public.

3 Signaling vice to signal authenticity

We argue that one particular form of authenticity signaling has become central to contemporary populist politics and deserves separate theoretical treatment: Vice signaling — the deliberate and public violation of established norms, including truth-telling, decorum, civility, and institutional respect, by leaders who use these transgressions to differentiate themselves from an establishment they portray as unauthentic and hypocritical.

3.1 High and low styles in politics

US President Donald Trump is a paradigmatic case of a politician whose appeal rests heavily on a refusal to perform conventional political virtue. At various times, Trump openly admitted not paying taxes¹, mocked disabled reporters², and veterans³, used crude language where decorum was expected⁴, and so forth. Italy’s Silvio Berlusconi is another example of a politician who openly flaunted established rules. His career was sustained against a staggering array of allegations that would have destroyed the career of almost any Western politician: Across four terms as Prime Minister, he faced allegations ranging from tax fraud to bribery and Mafia connections (Newell, 2018). Probably the most extreme contemporary case of vice signaling is Philippines’ former President Rodrigo Duterte, who openly supported extrajudicial killings and made graphic sexual jokes at press conferences, among other things (Heydarian, 2017), and regularly used profanities in public speeches (Montiel et al., 2022). Similar behaviors are characteristic of populist politicians in general; Moffitt (2016) points to “bad manners” as a key way they establish their credentials and separate themselves from competitors. Relatedly, Ostiguy (2009) argues for a separate “high-low” political dimension, contrasting the polished “high” style with the less inhibited “low” style espoused by populist politicians.

Our theory argues that vice signaling confirms a politician’s legitimacy as someone who, even being flawed, does not hide one’s true intentions. Authenticity signaling is argued to be at the center of Donald Trump’s disruptive and norm-shattering rhetoric

¹See, e.g., “Trump brags about not paying taxes: ‘That makes me smart’”, *CNBC*, September 26, 2016.

²See, e.g., “Donald Trump criticised for mocking reporter with disability”, *BBC*, November 26, 2015.

³See, e.g., “Veterans Slam Trump For Reportedly Disparaging Military Service Members”, *Boston Globe*, September 4, 2020.

⁴See, e.g., “The Profanity President: Trump’s Four-Letter Vocabulary”, *The New York Times*, May 19, 2019 or “Profanity is a hallmark of Trump’s second term, a Post analysis finds”, *Washington Post*, May 21, 2026.

(Jamieson and Taussig, 2017; Litvin, 2023). Hahl et al. (2018) provide survey evidence showing that Trump’s supporters understood that some of his public statements were false, yet perceived his willingness to make conspicuously false statements and break with the norms as a sign of authenticity. Silvio Berlusconi’s legal troubles were likewise constituent parts of his public identity rather than embarrassments to be managed, contributing to his constructed image as a flawed, but authentic politician (Bennister and Worthy, 2024). Similarly, Boris Johnson’s unconventional image was cultivated to signal the willingness to defy conventional manners and ethics (Margulies, 2019; Bennister and Worthy, 2024). Importantly for our argument, displays of authenticity, while constituting strategic performances to attract voter support, are meant to be perceived as non-strategic and non-performative (Lacatus and Meibauer, 2022); in our model, vice signaling in equilibrium serves to communicate that the candidate is nonstrategic and therefore incapable of hiding one’s true self.

3.2 Political cynicism invites vice signaling

Besides the sender’s authenticity, vice signaling conveys two interlocking pieces of information about the competition. The first one is that their civic appearance is only an appearance; at the bottom, they are as self-interested and immoral as the vice-signaling candidate. Second, their polish and moral language also makes them guilty of being inauthentic hypocrites, while the vice-signaling candidate is the only one who is willing to acknowledge one’s inevitably flawed nature.

Our results show that this part of “bad but authentic” appeal is especially effective with politically cynical voters who view politicians as motivated primarily by self-interest (Cappella and Jamieson, 1997; Roudakova, 2017; Neumann and Zaki, 2023). Inherent cynical attitudes are particularly likely to be manifested when witnessing malfeasance (Altay et al., 2025; Lee et al., 2023), leading to sweeping and deterministic conclusions about the corrupt nature of politicians. This way, cynicism has been characterized as a mental shortcut enabling one to quickly process political information (Bartlett, 1995; Shields, 2021). Our model predicts that vice signaling will have an additional effect on cynical voters, portraying the competitor as someone who is both immoral and is doubly dangerous because the capacity to hide one’s bad intentions (Cappella and Jamieson, 1997). This dynamic was empirically demonstrated with a recent survey of Kenyan voters (Palmer and McClendon, 2026); in a survey vignette, cynical voters were more likely to favor a candidate who was caught stealing and accused his competitor of also being corrupt and only appearing clear, even though the voters viewed corruption as an undesirable trait. Similarly, Silvio Berlusconi owed his resilience to the deep-rooted cynicism of the public and popular perceptions of official hypocrisy; new allegations only served

to reinforce the public’s beliefs in the corrupt nature of all politicians (Newell, 2018). In our model, we show that political cynicism will lead to more vice signaling by politicians, increasing the probability that bad candidates are elected. This way, we expect vice-signaling to correlate with the predictors of political cynicism such as an authoritarian past (Zhelnina, 2020) or poor institutional environment (Stavrova and Ehlebracht, 2019).

4 Analysis

4.1 The model

There are two candidates competing for office, and a single representative voter.

Candidates. Each candidate $i \in \{1, 2\}$ is defined by two privately observed characteristics. First is the candidate’s *integrity* or underlying moral character $m_i \in \{H, L\}$. The value of $m_i = H$ corresponds to a high-integrity or virtuous candidate, while $m_i = L$ corresponds to a low-integrity or vicious one.

Let $p = \Pr(m_i = H) \in (0, 1)$ denote the prior probability that a candidate is high-integrity. We also assume that the integrity types of candidates are potentially correlated with the coefficient $\rho \in [0, 1)$. This can account for the voter’s belief that both candidates are drawn from the same pool; hence, one candidate’s moral character can be informative about the moral character of the competitor.

Second, candidates vary by their *authenticity* $a_i \in \{A, D\}$, with $a_i = A$ denoting an authentic candidate and $a_i = D$ as disingenuous. Let $\tau = \Pr(a_i = A) \in (0, 1)$ denote the prior probability that a candidate is authentic. We assume that authenticity types are independent across candidates and of their morality types.

After observing his type, each candidate chooses to present his moral character to the public. Let $s_i \in \{H, L\}$ denote the candidate’s style of presentation, with $s_i = H$ denoting the candidate’s decision to present himself as virtuous, high-integrity type, while $s_i = L$ denoting the decision to signal vice or appear as a low-integrity type.

Authentic candidates are assumed to be non-strategic. Instead, they always truthfully communicate their moral type: $s_i = m_i$. This operationalizes our notion of authenticity as the valued inability to misrepresent one’s true moral self. Disingenuous candidate i is strategic and can choose any s_i .

The voter. There is a single voter who observes candidates’ presentational styles (s_1, s_2) but not their morality or authenticity types. After observing the candidates, the voter decides what candidate to support; let $v \in \{1, 2\}$ denote the voter’s choice.

Payoffs. From a voter's perspective, both authenticity and moral character are desired properties of a candidate. The candidate $v \in \{1, 2\}$ whom the voter supports is elected, yielding the following payoff to the voter:

$$U(v) = m_v + \alpha a_v + \varepsilon_v, \quad (1)$$

where $\alpha > 0$ is how much the voter values authenticity of the elected candidate, relative to that candidate's moral character. The value ε_v is the extra utility that the voter receives when candidate v is elected, with $\varepsilon_2 = 0$ and $\varepsilon_1 \sim U\left[-\frac{1}{2\psi}, \frac{1}{2\psi}\right]$ is the utility shock observed by the voter but not by the candidates, with $0 < \psi \leq \frac{1}{4(1+\alpha)}$.

The candidates value getting elected. In addition, we assume that vice-signaling imposes a direct cost on the candidates. The vice signaler closes off paths to elite respectability and coalition partnership and exposes himself to additional legal and reputational risks. This translates to the following payoff function for candidate $i \in \{1, 2\}$:

$$u_i = \mathbb{1}_{i=v} - c(m_i, s_i), \quad (2)$$

where

$$c(m_i, s_i) = \begin{cases} 0 & \text{if } s_i = H, \\ c_H & \text{if } s_i = L, m_i = H \\ c_L & \text{if } s_i = L, m_i = L, \end{cases} \quad (3)$$

with

$$c_H > c_L \geq 0.$$

The final inequality reflects our assumption that for a high-integrity candidate, vice signaling is more expensive because it is more uncomfortable, more visible as performance, more difficult to sustain, and more damaging to the candidate's authentic interests.

4.2 Equilibrium

Our solution concept is the Bayesian equilibrium. Denote by $(\beta_{1H}, \beta_{1L}, \beta_{2H}, \beta_{2L})$ the strategy profile of the candidates, with β_{im} denoting the probability that disingenuous candidate $i \in \{1, 2\}$ will choose to vice signal if that candidate's moral type is $m \in \{H, L\}$.

Candidate strategies define the ex-post beliefs of the voter about the types of the candidates, conditional on how the candidates choose to present themselves. Define

$$\mu_i^m(s_1, s_2) = \Pr(m_i = H \mid s_1, s_2), \quad \mu_i^a(s_1, s_2) = \Pr(a_i = A \mid s_1, s_2) \quad (4)$$

the posterior beliefs about candidate i 's morality and authenticity, respectively. Since every candidate is authentic with a positive probability, these beliefs can be derived via the Bayes rule from any strategy profile of the disingenuous candidates.

We can immediately write the voter's expected utility of supporting candidate $v \in \{1, 2\}$, given how the candidates present themselves:

$$U(v|(s_1, s_2)) = \mu_v^m(s_1, s_2) + \alpha \mu_v^a(s_1, s_2) + \varepsilon_v. \quad (5)$$

Putting

$$\Delta_m(s_1, s_2) = \mu_1^m(s_1, s_2) - \mu_2^m(s_1, s_2), \quad \Delta_a(s_1, s_2) = \mu_1^a(s_1, s_2) - \mu_2^a(s_1, s_2), \quad (6)$$

we can write out the probability that candidate 1 wins:

$$P(s_1, s_2) = \Pr(U(1) \geq U(2)) = \frac{1}{2} + \psi(\Delta_m(s_1, s_2) + \alpha \Delta_a(s_1, s_2)). \quad (7)$$

This, in turn, allows us to write out the expected utilities of the candidates as

$$u_1(s_1, s_2) = P(s_1, s_2) - c(m_1, s_1), \quad u_2(s_1, s_2) = 1 - P(s_1, s_2) - c(m_2, s_2). \quad (8)$$

As a result, the first stage of this game becomes a psychological game where observer beliefs directly enter into the utility functions of the players (Battigalli and Dufwenberg, 2022).

We focus on the equilibria that are symmetric, meaning that $\beta_{1H} = \beta_{2H}$ and $\beta_{1L} = \beta_{2L}$. Our first result establishes the following:

Proposition 1. There exists a unique symmetric equilibrium with $\beta_L < 1$ and $\beta_H = 0$.

This result captures the central logic of vice signaling in our model. In order for vice signaling to be a viable strategy, there must be some low-integrity candidates who pretend to be high-integrity.⁵

Because vice-signaling is more costly for high-integrity candidates, high-integrity vice-signaling can occur only when the electoral gain from vice-signaling is sufficiently larger for high-integrity types than for low-integrity types. A high-integrity type can vice-signal only if the low-integrity type vice-signals with probability one, which rules

⁵Since the candidates are ex ante identical, with the same type distributions, signaling opportunities, and payoff functions, the natural benchmark is a symmetric equilibrium. Any asymmetry between candidate strategies would necessarily rely on arbitrary expectations about the candidates that are unrelated to the model's fundamentals. Moreover, numeric simulations suggest that asymmetric equilibria are unlikely, if ever possible.

out vice signaling by high-integrity candidates.⁶

As high-integrity candidates do not vice-signal in equilibrium, a candidate who chooses to vice-signal is identified by voters as low-integrity. When facing two candidates, one of whom is presenting himself as high-integrity type and the other as low-integrity, the latter appears more authentic. To the voter, such a candidate might be either a low-integrity authentic type, or a low-integrity inauthentic type who has chosen to vice-signal. A candidate who presents himself as high-integrity, instead, is either a high-integrity type, or someone who is both low-integrity and inauthentic. In case there is vice signaling in equilibrium, the low-integrity type is indifferent and randomizes between these two options. Alternatively, there can be no vice signaling by either low or high integrity types: The reputation benefit of revealing one's low integrity type is too small relative to the cost of doing so.

The following condition characterizes when the equilibrium involves no strategic vice-signaling:

Proposition 2. In the equilibrium, we have $\beta_L = 0$ if and only if

$$\frac{\alpha(1-\tau)-p(1-\rho)}{1-\tau[(1-p)+\rho p]} \leq \frac{c_L}{\psi}. \quad (9)$$

We will next examine how vice signaling is affected by the fundamentals of our model: The frequency of the authentic and moral types, costs of appearing as low-integrity type, the value that the voter places on authenticity, and how uncertain the voter's preferences are.

Symmetric-equilibrium uniqueness follows because, after establishing that every such equilibrium satisfies $\beta_H = 0$, the only remaining strategic object is the probability with which low-integrity candidates send the vice signal. The electoral gain from vice-signaling is strictly decreasing in this probability. When vice-signaling is rare, it is highly informative about authenticity and therefore electorally valuable. As vice-signaling becomes more common, it loses informational content and its electoral value falls. Consequently, the benefit from vice-signaling can cross the cost threshold at most once, yielding a unique symmetric equilibrium.

We proceed to define the conditions under which vice signaling is possible, and comparative statics. The following is true:

⁶For some ranges of parameter values — in particular, when there is no or weak correlation between candidate morality types — we can also completely rule out equilibria (symmetric or not) where the high-integrity type vice signals.

Proposition 3. We have $\beta_L > 0$ if and only if

$$\frac{\alpha(1-\tau)-p(1-\rho)}{1-\tau[(1-p)+\rho p]} > \frac{c_L}{\psi}. \quad (10)$$

The right-hand side of (10) is equal to the cost of vice signaling relative to the responsiveness of voters to signaling; if this value is lower, vice signaling becomes possible under a broader set of other parameters. Hence, it is useful to first consider a case where c_L is close to zero and political competition becomes a cheap talk game.

Even under a cheap talk game, vice signaling will not be possible if the voters do not value authenticity, or if authentic candidates are sufficiently common and τ is close to one. Authenticity must be sufficiently valued and sufficiently rare for candidates to consider revealing their immoral type even if there are no direct costs involved.

At the same time, vice signaling in a cheap talk game is possible if one of two conditions is satisfied. First, the share p of high-integrity types can be sufficiently low; second, ρ can be close to 1 so the public is sufficiently cynical. Either condition reduces the penalty of revealing one's vice to the voters. In the first case, the public does not believe that any candidate is a high-integrity type. In the second case, because morality types are correlated across candidates, the act of signaling vice damages the reputation of a competitor who professes to be virtuous; this is what we should expect with politically cynical voters, when malfeasance by one candidate leads the voters to make sweeping generalizations that apply to everyone.

Similar considerations are present in the more generic case of costly vice signaling: The left-hand side of (10) increases in α and ρ and decreases in p and τ . In case β_L is positive, its comparative statics are largely similar:

Proposition 4. Suppose that condition (10) is satisfied. Then β_L increases in α , ρ , and ψ , and decreases in p and c_L .

The effect of τ on β_L is, however, conditional. On one hand, a higher share of authentic types makes vice-signaling more effective, because an openly immoral candidate is more likely to be authentic. On the other hand, a higher share of authentic types leads to a reduction in the need for disingenuous immoral candidates to randomize. More formally, the effect of τ is positive whenever the increase in the authenticity premium generated by a larger share of authentic candidates outweighs the mechanical reduction in the need for disingenuous low-integrity candidates to randomize. Since neither of these two effects dominates globally, the overall effect of τ is conditional.

Our final set of results is concerned with how vice signaling affects the outcomes of elections. We first focus on the integrity of the elected candidates. In particular, how likely is a low-integrity type to defeat a high-integrity type? We have the following

result:

Proposition 5. If $\beta_L = 0$, we have

$$\Pr(1 \text{ wins} | m_1 = H, m_2 = L) = \frac{1}{2} - \frac{\psi \tau (\alpha(1 - \tau) - p(1 - \rho))}{1 - \tau(1 - p + p\rho)}. \quad (11)$$

This value is decreasing in α and ρ , and is increasing in p .

If $\beta_L > 0$, then

$$\Pr(1 \text{ wins} | m_1 = H, m_2 = L) = \frac{1}{2} - c_L x, \quad (12)$$

where $x = \tau + (1 - \tau)\beta_L$ is the probability that a low-integrity candidate signals L . This value is weakly decreasing in α , τ , ρ , and ψ , decreasing in c_L , and increasing in p .

In the case without vice signaling, the only types that send the vice signal are authentic low-integrity types, and observing $s_i = L$ necessarily implies that candidate i is authentic. Hence, the probability that a high-integrity candidate is elected declines with α or how much the voters value authenticity. If the share p of high-integrity candidates is larger, then a high-integrity type is more likely to stand behind a signal $s_i = H$, so the probability that a high-integrity candidate is elected increases, conditional on candidates sending different signals. Higher ρ , on the contrary, means that a candidate who signals L casts a shadow over someone signaling H , leading to a high-integrity candidate less likely to be selected.

In the case where vice signaling is present, it is done by low-integrity candidates and therefore reduces the probability that the high-integrity candidate defeats a low-integrity opponent. The probability that the high-integrity candidate wins is $\frac{1}{2}$ minus the cost c_L times the ex ante probability that a low-integrity candidate signals vice. As a result, higher public cynicism, greater value the public puts on authenticity, and a higher share of authentic candidates all lead to low-integrity candidates being elected more frequently.

A higher cost c_L directly discourages vice-signaling and lowers $\tau + (1 - \tau)\beta_L$, but in a mixed equilibrium the marginal benefit from vice-signaling must equal c_L . This raises the payoff advantage associated with the vice-signaling states strongly enough to increase the immoral candidate's equilibrium winning probability against a moral opponent. This way, raising the reputational cost of transgression doesn't shield voters from bad candidates. Although it reduced the probability of vice signaling, it simultaneously makes the vice signal more informative, making the vice-signaling candidate more likely to win.

Proposition 5 analyzed the probability that a high-integrity candidate is elected; hence, it focused on the case where candidates of different integrity levels competed against each other. However, we are also assuming that voters value authenticity, so, in order to study the effects of vice signaling on voter welfare, we need to consider the

effect of vice signaling on the authenticity of elected candidates. Vice signaling can affect that probability even when two low-integrity candidates are competing. Hence, we look at the *ex ante* expected voter welfare:

$$W^Q = \sum_{s_1, s_2} \Gamma_{s_1 s_2} [P(s_1, s_2) v_1(s_1, s_2) + (1 - P(s_1, s_2)) v_2(s_1, s_2)]. \quad (13)$$

where $v_i(s_1, s_2) := \mu_i^m(s_1, s_2) + \alpha \mu_i^a(s_1, s_2)$ and $\Gamma_{s_1 s_2}$ is the probability of signal profile (s_1, s_2) given equilibrium strategies.

The following result summarizes our analysis of expected voter welfare:

Proposition 6. If $\beta_L = 0$, we have

$$W^Q = p + \alpha\tau + \frac{2\psi\tau(1-p)(\alpha(1-\tau) - p(1-\rho))^2}{1 - \tau(1 - p + p\rho)}. \quad (14)$$

If $\beta_L > 0$, we have

$$W^Q = p + \alpha\tau + \frac{2c_L^2(1-p)}{\psi} x(1 - (1 - p + p\rho)x). \quad (15)$$

The value W^Q is strictly increasing in c_L when $\beta_L > 0$ and $c_L > 0$.

The first component of (15) and (14), $p + \alpha\tau$, is the expected *ex ante* quality of a candidate. The second component is the value of information generated by campaign signals: it captures how much voters gain from selecting between candidates on the basis of posterior differences in morality and authenticity. The value of information is always positive in our setting, as the single voter is always free to ignore candidate signals and choose a candidate at random, receiving the expected payoff $p + \alpha\tau$.

When $\beta_L = 0$, vice signals are sent only by authentic low-integrity candidates, so observing L perfectly reveals authenticity and observing H increases the posterior probability that a candidate is high-integrity. The effects of parameters depend on whether they strengthen or weaken the informational distinction created by the signals.

When $\beta_L > 0$, low-integrity disingenuous candidates sometimes imitate authentic candidates by vice-signaling. Parameters then affect welfare through two channels. On the one hand, they change the informational content of signals. On the other hand, they change the equilibrium frequency of vice-signaling. Welfare increases when the first effect dominates the second, that is, when a parameter improves voter inference more than it encourages strategic imitation. Conversely, welfare falls when the increase in vice-signaling erodes the informational value of observed signals.

For c_L , the information effect dominates in the mixed equilibrium. A higher c_L lowers the equilibrium frequency of strategic vice-signaling, making observed vice signals more

informative about authenticity. Although this can reduce the probability that a high-integrity candidate defeats a low-integrity opponent, voters are compensated through the authenticity channel, and the total welfare expression W^Q increases with c_L .

5 Conclusion

Our paper offers an account why bad behavior of politicians is advertised and rewarded — a fact even more surprising as transgressive political actors face substantial costs, such as the explicit refusal of other political actors to cooperate with transgressive actors (Van Spanje and Van Der Brug, 2007).

A conventional account is that such behavior communicates the closeness of the politician to the voter, bridging the symbolic gap between them and, at the same time, separating the politician from the vilified “elite” (Moffitt, 2016). Our analysis points to a further mechanism that operates alongside this one. The open display of one’s flaws is not only a claim of closeness but a credible signal of *authenticity* — a valued property conceptualized as the inability of strategically manipulating one’s public self. This assumption is consistent with the account of populist politicians trying to appear spontaneous, impulsive, or unsophisticated. At the same time, we are conventionally maintaining that, *ceteris paribus*, voters prefer high-integrity candidates; this way, we abstract from the possibility that corrupt practices are normalized (Kubbe et al., 2024) and, for example, corrupt or morally deficient candidates are perceived as more efficient.

In a contested race this signal does double duty. It authenticates the candidate who sends it, and at the same time, “unmasks” a rival who professes virtue, casting them as a dishonest hypocrite. The latter effect is stronger if the voters are more cynical; hence, one of the predictions of our analysis is that we should observe more bad behavior performed in front of a more cynical public.⁷

Another result concerns the effect of punishing transgression on the electoral success of transgressive candidates. Legal and criminal sanctions for the underlying conduct, the threat of losing office eligibility, gatekeeping within own political parties, refusal of other parties to cooperate or form coalitions, withdrawal of sponsor support, social and elite ostracism — all of those are very tangible costs that may reduce the incidence of vice signaling, yet make it more informative to the voters; in our model, this ultimately leads to more bad candidates being elected. Some historical efforts to exact electoral accountability have been successful (Ferraz and Finan, 2008), but we show that their effect can be reduced, or even reversed, in the presence of vice signaling.

⁷In a recent paper, Zakharov (2026) similarly argued that media outlets will be more likely to bias their reporting if the audiences are inherently more cynical, because in such audiences a conspicuously false message will reduce trust in the competing news sources.

In future work, our setting can be extended in several ways. One intriguing possibility would be to model how the public's cynicism preferences for authenticity are shaped by their previous experience with electing candidates who follow different political styles. Further work can identify path dependencies, when cynical beliefs and the demand for authenticity become self-reinforcing. Another option is to consider multiple audiences that vary in ex ante beliefs, affinities, or attunement to different candidates; we posit that performative populism can be more pronounced in the presence of political polarization. Finally, strategic intermediaries —journalists, social-media platforms, or party elites — can curate, amplify, or contest candidate signals, reshaping both the candidates' incentive to signal vice and the audience's response.

References

- Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin**, "A political theory of populism," *The quarterly journal of economics*, 2013, 128 (2), 771–805.
- Aiolfi, Théo**, "Populism as a transgressive style," *Global Studies Quarterly*, 2022, 2 (1), ksac006.
- Altay, Sacha, Benjamin A Lyons, and Ariana Modirrousta-Galian**, "Exposure to higher rates of false news erodes media trust and fuels overconfidence," *Mass Communication and Society*, 2025, 28 (2), 301–325.
- Bailey, Erica R and Sheena S Iyengar**, "Yours truly: On the complex relationship between authenticity and honesty," *Current opinion in psychology*, 2022, 47, 101419.
- Bartlett, Frederic Charles**, *Remembering: A study in experimental and social psychology*, Cambridge university press, 1995.
- Battigalli, Pierpaolo and Martin Dufwenberg**, "Belief-dependent motivations and psychological game theory," *Journal of Economic Literature*, 2022, 60 (3), 833–882.
- Bennister, Mark and Ben Worthy**, "The 'electoral presidentialization' of Silvio Berlusconi and Boris Johnson: Chaos, controversy, and lost chances," *The British Journal of Politics and International Relations*, 2024, 26 (4), 1231–1251.
- Bhatti, Yosef, Kasper M Hansen, and Asmus Leth Olsen**, "Political hypocrisy: The effect of political scandals on candidate evaluations," *Acta Politica*, 2013, 48 (4), 408–428.
- Bobba, Giuliano and Nicolas Hubé**, *Populism and the Politicization of the COVID-19 Crisis in Europe*, Springer Nature, 2021.

- Buisseret, Peter and Carlo Prato**, “Electoral control and the human capital of politicians,” *Games and Economic Behavior*, 2016, 98, 34–55.
- Cappella, Joseph N and Kathleen Hall Jamieson**, *Spiral of cynicism: The press and the public good*, Oxford University Press on Demand, 1997.
- Caselli, Francesco and Massimo Morelli**, “Bad politicians,” *Journal of public economics*, 2004, 88 (3-4), 759–782.
- Effron, Daniel A, Hazel Rose Markus, Lauren M Jackman, Yukiko Muramoto, and Hamdi Muluk**, “Hypocrisy and culture: Failing to practice what you preach receives harsher interpersonal reactions in independent (vs. interdependent) cultures,” *Journal of Experimental Social Psychology*, 2018, 76, 371–384.
- , **Kieran O’Connor, Hannes Leroy, and Brian J Lucas**, “From inconsistency to hypocrisy: When does “saying one thing but doing another” invite condemnation?,” *Research in Organizational Behavior*, 2018, 38, 61–75.
- Enli, Gunn**, “Twitter as arena for the authentic outsider: Exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election,” *European journal of communication*, 2017, 32 (1), 50–61.
- , “Populism as “truth”: How mediated authenticity strengthens the populist message,” *The International Journal of Press/Politics*, 2025, 30 (1), 83–99.
- Fancourt, Daisy, Andrew Steptoe, and Liam Wright**, “The Cummings effect: politics, trust, and behaviours during the COVID-19 pandemic,” *The lancet*, 2020, 396 (10249), 464–465.
- Ferraz, Claudio and Frederico Finan**, “Exposing corrupt politicians: the effects of Brazil’s publicly released audits on electoral outcomes,” *The Quarterly journal of economics*, 2008, 123 (2), 703–745.
- Fieschi, Catherine**, *Populocracy: The tyranny of authenticity and the rise of populism*, Newcastle: Agenda Publishing, 2019.
- Gailmard, Lindsey**, “Electoral accountability and political competence,” *Journal of Theoretical Politics*, 2022, 34 (2), 236–261.
- Gunster, Shane, Darren Fleet, Matthew Paterson, and Paul Saurette**, ““Why don’t you act like you believe it?”: competing visions of climate hypocrisy,” *Frontiers in communication*, 2018, 3, 49.

- Guriey, Sergei and Elias Papaioannou**, “The political economy of populism,” *Journal of Economic Literature*, 2022, 60 (3), 753–832.
- Hahl, Oliver, Minjae Kim, and Ezra W Zuckerman Sivan**, “The authentic appeal of the lying demagogue: Proclaiming the deeper truth about political illegitimacy,” *American Sociological Review*, 2018, 83 (1), 1–33.
- Hawkins, Kirk A**, “Is Chávez populist? Measuring populist discourse in comparative perspective,” *Comparative political studies*, 2009, 42 (8), 1040–1067.
- Heydarian, Richard Javad**, *The rise of Duterte: A populist revolt against elite democracy*, Springer, 2017.
- Hunger, Sophia**, “Virtuous people and evil elites? The role of moralizing frames and normative distinctions in identifying populist discourse,” *Political Research Exchange*, 2024, 6 (1), 2370308.
- Huttunen, Kiia Jasmin Alexandra and Stephan Lewandowsky**, “Tolerance for democratic norm violations increases when sincerity replaces accuracy as a marker of honesty,” *Communications Psychology*, 2026.
- Jamieson, Kathleen Hall and Doron Taussig**, “Disruption, demonization, deliverance, and norm destruction: The rhetorical signature of Donald J. Trump,” *Political Science Quarterly*, 2017, 132 (4), 619–650.
- Jordan, Jillian and Roseanna Sommers**, “When does moral engagement risk triggering a hypocrite penalty?,” *Current Opinion in Psychology*, 2022, 47, 101404.
- Jordan, Jillian J, Roseanna Sommers, Paul Bloom, and David G Rand**, “Why do we hate hypocrites? Evidence for a theory of false signaling,” *Psychological science*, 2017, 28 (3), 356–368.
- Kakutani, Shizuo**, “A generalization of Brouwer’s fixed point theorem,” *Duke Mathematical Journal*, 1941, 8 (3), 457–459.
- Kartik, Navin and Richard Van Weelden**, “Informative cheap talk in elections,” *The Review of Economic Studies*, 2019, 86 (2), 755–784.
- Kazin, Michael**, *The populist persuasion: An American history*, Cornell University Press, 1998.
- Kubbe, Ina, Claudia Baez-Camargo, and Cheyanne Scharbatke-Church**, “Corruption and social norms: A new arrow in the quiver,” *Annual Review of Political Science*, 2024, 27 (1), 423–444.

- Lacatus, Corina and Gustav Meibauer**, ““Saying it like it is”: Right-wing populism, international politics, and the performance of authenticity,” *The British Journal of Politics and International Relations*, 2022, 24 (3), 437–457.
- Lee, Sangwon, Homero Gil de Zúñiga, and Kevin Munger**, “Antecedents and consequences of fake news exposure: A two-panel study on how news use and different indicators of fake news exposure affect media trust,” *Human Communication Research*, 2023, 49 (4), 408–420.
- Litvin, Boris**, ““This Hearing Should Be Flipped”: Democratic Spectatorship, Social Media, and the Problem of Demagogic Candor,” *American Political Science Review*, 2023, 117 (1), 153–167.
- Luebke, Simon M**, “Political authenticity: Conceptualization of a popular term,” *The International Journal of Press/Politics*, 2021, 26 (3), 635–653.
- **and Dennis Steffan**, “Who values political authenticity? The antecedents and electoral consequences of voters’ preference for authentic politicians,” *Polity*, 2025, 57 (2), 328–349.
- **and Ines Engelmann**, ““Vote for Me, I Am Authentic”: Performed Political Authenticity on Social Media,” *Media and Communication*, 2026, 14.
- Mansbridge, Jane and Stephen Macedo**, “Populism and democratic theory,” *Annual review of law and social science*, 2019, 15 (1), 59–77.
- Margulies, Ben**, “Understanding Boris Johnson’s appeal: Performative rebelliousness, impishness, and indiscipline,” *LSE Policy, and Politics Blog*, 2019, 1.
- Mizuno, Nobuhiro and Ryosuke Okazawa**, “Why do voters elect less qualified candidates?,” *Journal of Theoretical Politics*, 2022, 34 (3), 443–477.
- Moffitt, Benjamin**, *The global rise of populism: Performance, political style, and representation*, Stanford University Press, 2016.
- Montiel, Cristina Jayme, Joshua Uyheng, and Nmanuel De Leon**, “Presidential profanity in duterte’s Philippines: How swearing discursively constructs a populist regime,” *Journal of Language and Social Psychology*, 2022, 41 (4), 428–449.
- Mudde, Cas**, *The far right today*, John Wiley & Sons, 2019.
- **and Cristóbal Rovira Kaltwasser**, *Populism: A very short introduction*, oxford university Press, 2017.

- Muller, Jan-Werner**, *What Is Populism?*, University of Pennsylvania Press, 2016.
- Nadesan, Majia**, “Crises narratives defining the COVID-19 pandemic: Expert uncertainties and conspiratorial sensemaking,” *American Behavioral Scientist*, 2022, p. 00027642221085893.
- Neumann, Eric and Jamil Zaki**, “Toward a social psychology of cynicism,” *Trends in Cognitive Sciences*, 2023, 27 (1), 1–3.
- Newell, James L**, “Silvio Berlusconi: A study in failure,” in “Silvio Berlusconi,” Manchester University Press, 2018.
- Ostiguy, Pierre**, “The high and the low in politics: a two-dimensional political space for comparative analysis and electoral studies,” *Working paper*, 2009.
- Padró i Miquel, Gerard and Erik Snowberg**, “The lesser evil: Executive accountability with partisan supporters,” *Journal of Theoretical Politics*, 2012, 24 (1), 19–45.
- Palmer, Alexis and Gwyneth McClendon**, “Cynicism and Voter Support for Openly Corrupt Candidates: Evidence from Kenya,” *Working paper*, 2026.
- Reuter, Ora John and David Szakonyi**, “Electoral manipulation and regime support: Survey evidence from Russia,” *World Politics*, 2021, 73 (2), 275–314.
- Rooduijn, Matthijs**, “The nucleus of populism: In search of the lowest common denominator,” *Government and opposition*, 2014, 49 (4), 573–599.
- Roudakova, Natalia**, *Losing Pravda: Ethics and the press in post-truth Russia*, Cambridge University Press, 2017.
- Schnakenberg, Keith**, “Candidate traits in elections: when good news for selection is bad news for accountability,” *Political Science Research and Methods*, 2021, 9 (1), 206–214.
- Shields, Paul**, “Killing politics softly: Unconvincing propaganda and political cynicism in Russia,” *Communist and Post-Communist Studies*, 2021, 54 (4), 54–73.
- Sorensen, Lone**, “Disrupting democracy: Democratization conflicts as performative struggles,” *Media, War & Conflict*, 2020, 13 (1), 8–26.
- Spanje, Joost Van and Wouter Van Der Brug**, “The party as pariah: The exclusion of anti-immigration parties and its effect on their ideological positions,” *West European Politics*, 2007, 30 (5), 1022–1040.

- Stavrova, Olga and Daniel Ehlebracht**, “The cynical genius illusion: Exploring and debunking lay beliefs about cynicism and competence,” *Personality and Social Psychology Bulletin*, 2019, 45 (2), 254–269.
- Strayhorn, Joshua A**, “Making the other side mad: How out-group distaste benefits less competent candidates,” *American Journal of Political Science*, 2026, 70 (1), 334–347.
- Valgarðsson, Viktor, Will Jennings, Gerry Stoker, Lawrence McKay, Daniel Devine, and Nick Clarke**, “The good politician: Competence, integrity and authenticity in seven democracies,” *Political Studies*, 2025, 73 (2), 631–656.
- Vries, Catherine E De and Hector Solaz**, “The electoral consequences of corruption,” *Annual Review of Political Science*, 2017, 20, 391–408.
- Winters, Matthew S and Rebecca Weitz-Shapiro**, “Lacking information or condoning corruption: When do voters support corrupt politicians?,” *Comparative Politics*, 2013, 45 (4), 418–436.
- Zakharov, Alexei**, “Propaganda to a cynical audience,” *Political Science Research and Methods*, 2026, pp. 1–12.
- Zheltnina, Anna**, “The apathy syndrome: How we are trained not to care about politics,” *Social Problems*, 2020, 67 (2), 358–378.

Appendix

A Proofs of Statements

Proof of Proposition 1. Consider symmetric equilibria, so that $\beta_{1H} = \beta_{2H} \equiv \beta_H$ and $\beta_{1L} = \beta_{2L} \equiv \beta_L$.

First, we show existence. Let $K := [0, 1]^2$ be the symmetric strategy space, where $\beta = (\beta_H, \beta_L)$ specifies the probabilities with which disingenuous high- and low-integrity types vice-signal. Given any $\beta \in K$, Bayes' rule determines voter beliefs after every positive-probability signal profile. For each type $m \in \{H, L\}$, let $D_m(\beta)$ be the net payoff gain from sending L rather than H . The functions $D_m(\beta)$ are continuous in β . Hence the best-response correspondence

$$BR_m(\beta) = \begin{cases} \{1\}, & D_m(\beta) > 0, \\ [0, 1], & D_m(\beta) = 0, \\ \{0\}, & D_m(\beta) < 0 \end{cases}$$

is nonempty, convex-valued, closed-valued, and upper hemicontinuous. Therefore $BR(\beta) := BR_H(\beta) \times BR_L(\beta)$ maps the compact convex set K into itself and satisfies the hypotheses of **Kakutani's (1941)** fixed-point theorem. Hence there exists a symmetric equilibrium.

Next, in any symmetric strategy profile, when both candidates send the same signal, they have the same posterior voter value, so $P(H, H) = P(L, L) = \frac{1}{2}$. Symmetry also implies $P(L, H) = 1 - P(H, L)$. Therefore, for a disingenuous type m , the payoff gain from sending L rather than H is independent of the opponent's signal and equals $\frac{1}{2} - P(H, L) - c_m$. Thus

$$\beta_m = \begin{cases} 1, & c_m < \frac{1}{2} - P(H, L), \\ [0, 1], & c_m = \frac{1}{2} - P(H, L), \\ 0, & c_m > \frac{1}{2} - P(H, L). \end{cases}$$

Since both types face the same gross electoral gain from vice-signaling, but $c_H > c_L$, the best-response characterization above implies that $\beta_H > 0$ only if $\beta_L = 1$.

Suppose that $\beta_H > 0$ and $\beta_L = 1$. A high-integrity candidate sends L with probability $x_H = (1 - \tau)\beta_H < 1$, while a low-integrity candidate sends L with probability $x_L = \tau + (1 - \tau)\beta_L = 1$.

Put $\pi_{HH} = p^2 + \rho p(1 - p)$ and $\pi_{HL} = p(1 - p)(1 - \rho)$ the prior probabilities of the

morality profiles (H, H) and (H, L) . Under $\beta_L = 1$, a low-integrity candidate sends L with probability one. Hence, after the signal profile (H, L) , candidate 1 must be high-integrity. Candidate 2, however, need not be low-integrity, because a high-integrity disingenuous candidate sends L with probability $x_H > 0$.

Conditional on (H, L) , Bayes' rule places positive probability only on the morality profiles (H, H) and (H, L) . Bayes' rule gives $\mu_1^m(H, L) = 1$ and $\mu_2^m(H, L) = \frac{\pi_{HH}x_H}{\pi_{HH}x_H + \pi_{HL}}$. Therefore $\Delta_m(H, L) = 1 - \frac{\pi_{HH}x_H}{\pi_{HH}x_H + \pi_{HL}} = \frac{\pi_{HL}}{\pi_{HH}x_H + \pi_{HL}} > 0$. For authenticity, Bayes' rule gives $\mu_1^a(H, L) = \frac{\tau}{1-x_H}$, because a high-integrity candidate sends H with total probability $1 - x_H$, and the authentic high-integrity type contributes probability τ . For candidate 2, $\mu_2^a(H, L) = \frac{\tau\pi_{HL}}{\pi_{HH}x_H + \pi_{HL}}$, because candidate 2 is authentic and sends L only when he is low-integrity authentic; a high-integrity authentic candidate sends H . Hence $\Delta_a(H, L) = \frac{\tau}{1-x_H} - \frac{\tau\pi_{HL}}{\pi_{HH}x_H + \pi_{HL}} \geq 0$. Thus

$$P(H, L) = \frac{1}{2} + \psi \left[\frac{\pi_{HL}}{\pi_{HH}x_H + \pi_{HL}} + \alpha \left(\frac{\tau}{1-x_H} - \frac{\tau\pi_{HL}}{\pi_{HH}x_H + \pi_{HL}} \right) \right] > \frac{1}{2}.$$

So $\frac{1}{2} - P(H, L) - c_L < 0$, which means the low-integrity disingenuous type strictly prefers sending H to sending L . This contradicts $\beta_L = 1$. Combined with the implication $\beta_H > 0 \Rightarrow \beta_L = 1$ established above, it follows that $\beta_H = 0$ in every symmetric equilibrium.

It remains to show uniqueness. Let $x := \tau + (1 - \tau)\beta_L$ and $A := 1 - p + \rho p$. With $\beta_H = 0$, Bayes' rule gives the low-integrity disingenuous type's gross electoral gain from vice-signaling as $B(x) = \psi \frac{\alpha\tau(1-x) - p(1-\rho)x}{x(1-Ax)}$. A symmetric equilibrium requires the low-integrity disingenuous type to compare $B(x)$ with c_L .

Differentiating, $B'(x) = -\psi \frac{\alpha\tau[1-2Ax+Ax^2] + p(1-\rho)Ax^2}{x^2(1-Ax)^2}$. The denominator is positive. The numerator is positive because $1 - 2Ax + Ax^2 = (1 - Ax)^2 + A(1 - A)x^2 > 0$, and $p(1 - \rho)Ax^2 \geq 0$. Therefore $B'(x) < 0$.

If $B(\tau) \leq c_L$, then $\beta_L = 0$ is a symmetric equilibrium. Since B is strictly decreasing, no $x > \tau$ can satisfy $B(x) = c_L$, so this symmetric equilibrium is unique. If $B(\tau) > c_L$, then $\beta_L = 0$ is not an equilibrium. Since $B(1) = -\psi < c_L$ and B is continuous and strictly decreasing, there exists a unique $x^* \in (\tau, 1)$ such that $B(x^*) = c_L$. The corresponding equilibrium mixing probability is $\beta_L^* = \frac{x^* - \tau}{1 - \tau}$. Therefore there exists a unique symmetric equilibrium, and in it $\beta_H = 0$ and $\beta_L < 1$. $\square$

Proof of Proposition 2. By Proposition 1, the unique symmetric equilibrium satisfies $\beta_H = 0$ and is characterized by $x := \tau + (1 - \tau)\beta_L$.

As derived in the proof of Proposition 1, the gross gain from vice-signaling for a low-integrity disingenuous type is $B(x) = \psi \frac{\alpha\tau(1-x) - p(1-\rho)x}{x(1-(1-p+\rho p)x)}$. At $\beta_L = 0$, we have $x = \tau$, so $B(\tau) = \psi \frac{\alpha(1-\tau) - p(1-\rho)}{1-\tau[(1-p)+\rho p]}$. The no-vice-signaling profile is the symmetric equilibrium if and only if $B(\tau) \leq c_L$. Dividing by $\psi > 0$ gives $\frac{\alpha(1-\tau) - p(1-\rho)}{1-\tau[(1-p)+\rho p]} \leq \frac{c_L}{\psi}$. $\square$

Proof of Proposition 3. By the proof of Proposition 1, vice-signaling occurs in the unique symmetric equilibrium if and only if the no-vice-signaling profile is not an equilibrium. At the no-vice-signaling profile, $\beta_L = 0$, $x_L = \tau$. The electoral gain from a low-integrity disingenuous candidate deviating to vice-signaling is

$$B(\tau) = \psi \frac{\alpha(1-\tau) - p(1-\rho)}{1-\tau[(1-p) + \rho p]}.$$

By the best-response comparison derived in the proof of Proposition 1, the no-vice-signaling profile fails to be an equilibrium if and only if $B(\tau) > c_L$. Substituting the expression for $B(\tau)$ gives

$$\frac{\alpha(1-\tau) - p(1-\rho)}{1-\tau[(1-p) + \rho p]} > \frac{c_L}{\psi}.$$

Thus $\beta_L > 0$ if and only if condition (10) holds. $\square$

Proof of Proposition 4. Suppose condition (10) holds. Then the unique symmetric equilibrium has $\beta_L > 0$, so the low-integrity disingenuous type mixes. Let $x_L^* := \tau + (1-\tau)\beta_L$. By the proof of Proposition 1, x_L^* is characterized by $\Phi(x_L^*; \alpha, \tau, p, \rho, \psi, c_L) = 0$, where $\Phi(x_L; \alpha, \tau, p, \rho, \psi, c_L) := \psi \frac{\alpha\tau(1-x_L) - p(1-\rho)x_L}{x_L(1-Ax_L)} - c_L$, and $A := 1-p+\rho p$. The proof of Proposition 1 established that

$$\frac{\partial \Phi}{\partial x_L} < 0.$$

The relevant partial derivatives are

$$\frac{\partial \Phi}{\partial \alpha} = \psi \frac{\tau(1-x_L)}{x_L(1-Ax_L)} > 0,$$

$$\frac{\partial \Phi}{\partial \tau} = \psi \frac{\alpha(1-x_L)}{x_L(1-Ax_L)} > 0,$$

$$\frac{\partial \Phi}{\partial p} = -\psi \frac{(1-\rho)(1+\alpha\tau)(1-x_L)}{(1-Ax_L)^2} < 0,$$

$$\frac{\partial \Phi}{\partial \rho} = \psi \frac{p(1+\alpha\tau)(1-x_L)}{(1-Ax_L)^2} > 0,$$

$$\frac{\partial \Phi}{\partial \psi} = \frac{\alpha\tau(1-x_L) - p(1-\rho)x_L}{x_L(1-Ax_L)} = \frac{c_L}{\psi} \geq 0,$$

and

$$\frac{\partial \Phi}{\partial c_L} = -1.$$

By the implicit function theorem, $\frac{\partial x_L^*}{\partial \theta} = -\frac{\partial \Phi / \partial \theta}{\partial \Phi / \partial x_L}$. Since $\partial \Phi / \partial x_L < 0$, it follows that

$$\frac{\partial x_L^*}{\partial \alpha} > 0, \quad \frac{\partial x_L^*}{\partial p} < 0, \quad \frac{\partial x_L^*}{\partial \rho} > 0, \quad \frac{\partial x_L^*}{\partial c_L} < 0, \quad \frac{\partial x_L^*}{\partial \tau} > 0.$$

Moreover,

$$\frac{\partial x_L^*}{\partial \psi} > 0 \quad \text{if } c_L > 0,$$

and

$$\frac{\partial x_L^*}{\partial \psi} = 0 \quad \text{if } c_L = 0.$$

Finally, $\beta_L = \frac{x_L^* - \tau}{1 - \tau}$. For parameters other than τ , this transformation is strictly increasing in x_L^* . Hence β_L increases in α , ρ , and ψ , and decreases in p and c_L . $\square$

Proof of Proposition 5. Suppose first that $\beta_L = 0$. Then low-integrity candidates send L only when authentic, so $x_L = \tau$. Using the posterior beliefs derived in the proof of Proposition 1, evaluated at $x_L = \tau$, we obtain $P(L, L) = \frac{1}{2} - \frac{\psi(\alpha(1-\tau) - p(1-\rho))}{1 - \tau(1-p+p\rho)}$. Since a high-integrity candidate always sends H and a low-integrity candidate sends L with probability τ , $P(1 \text{ wins} \mid m_1 = H, m_2 = L) = \tau P(L, L) + (1 - \tau)\frac{1}{2} = \frac{1}{2} - \frac{\psi\tau(\alpha(1-\tau) - p(1-\rho))}{1 - \tau(1-p+p\rho)}$.

The comparative statics follow immediately from differentiation. Suppose $\beta_L > 0$. Then the equilibrium is mixed, and the low-integrity disingenuous type is indifferent between sending H and sending L . By Proposition 1, $\frac{1}{2} - P(L, L) = c_L$, so $P(L, L) = \frac{1}{2} - c_L$. Now condition on candidate 1 being high-integrity and candidate 2 being low-integrity. By Proposition 1, high-integrity candidates never vice-signal in the symmetric equilibrium, so candidate 1 sends H with probability one. Candidate 2 sends L with probability $x_L^* = \tau + (1 - \tau)\beta_L$, and sends H with probability $1 - x_L^*$.

Therefore, $P(1 \text{ wins} \mid m_1 = H, m_2 = L) = x_L^* P(L, L) + (1 - x_L^*) P(L, L)$. Since $P(L, L) = 1/2$, this becomes $P(1 \text{ wins} \mid m_1 = H, m_2 = L) = x_L^* (\frac{1}{2} - c_L) + (1 - x_L^*) \frac{1}{2}$. Simplifying, $P(1 \text{ wins} \mid m_1 = H, m_2 = L) = \frac{1}{2} - c_L x_L^*$. Using $x_L^* = \tau + (1 - \tau)\beta_L$, we obtain $P(1 \text{ wins} \mid m_1 = H, m_2 = L) = \frac{1}{2} - c_L(\tau + (1 - \tau)\beta_L)$.

Importantly, for any parameter $\theta \in \{\alpha, \tau, p, \rho, \psi\}$, $\frac{\partial}{\partial \theta} P(1 \text{ wins} \mid m_1 = H, m_2 = L) =$

$$-c_L \frac{\partial x_L^*}{\partial \theta}.$$

By Proposition 4, x_L^* increases in α , τ , ρ , and ψ , and decreases in p . Therefore $P(1 \text{ wins} \mid m_1 = H, m_2 = L)$ is decreasing in α , τ , ρ , and ψ , and increasing in p .

It remains to determine the effect of c_L . From the mixed-equilibrium condition, $c_L = \psi \frac{\alpha\tau(1-x_L^*)-p(1-\rho)x_L^*}{x_L^*(1-Ax_L^*)}$. Multiplying by x_L^* gives $c_L x_L^* = \psi \frac{\alpha\tau(1-x_L^*)-p(1-\rho)x_L^*}{1-Ax_L^*}$. Hence $P(1 \text{ wins} \mid m_1 = H, m_2 = L) = \frac{1}{2} - \psi \frac{\alpha\tau(1-x_L^*)-p(1-\rho)x_L^*}{1-Ax_L^*}$. Differentiating this expression with respect to x_L^* gives $\frac{\partial}{\partial x_L^*} P(1 \text{ wins} \mid m_1 = H, m_2 = L) = \psi \frac{p(1-\rho)(1+\alpha\tau)}{(1-Ax_L^*)^2} > 0$. By Proposition 4, $\frac{\partial x_L^*}{\partial c_L} < 0$. Therefore $\frac{\partial}{\partial c_L} P(1 \text{ wins} \mid m_1 = H, m_2 = L) < 0$. Thus the high-integrity candidate's probability of defeating a low-integrity opponent is decreasing in c_L . $\square$

Proof of Proposition 6. Let $v_i(s_1, s_2) := \mu_i^m(s_1, s_2) + \alpha\mu_i^a(s_1, s_2)$. By symmetry, when both candidates send the same signal, they have the same expected voter value. Hence these signal profiles do not affect selection. The only welfare gain from information comes from asymmetric signal profiles.

For the asymmetric signal profile (H, L) , define $D(H, L) := v_1(H, L) - v_2(H, L)$. Since $P(L, L) = \frac{1}{2} + \psi D(H, L)$, the expected voter value conditional on (H, L) is $P(L, L)v_1(H, L) + (1 - P(L, L))v_2(H, L) = \frac{1}{2}[v_1(H, L) + v_2(H, L)] + \psi D(H, L)^2$. The same contribution arises from the opposite asymmetric profile (L, H) . Therefore, $W^Q = p + \alpha\tau + 2\psi \Gamma_{HL} D(H, L)^2$.

By Proposition 1, $\beta_H = 0$. Let $x := \tau + (1 - \tau)\beta_L$ and $A := 1 - p + p\rho$. Then $\Gamma_{HL} = (1 - p)x(1 - Ax)$. Moreover, the proof of Proposition 1 gives $D(H, L) = \frac{p(1-\rho)x - \alpha\tau(1-x)}{x(1-Ax)}$. Hence $W^Q = p + \alpha\tau + 2\psi(1 - p)x(1 - Ax) \left[\frac{p(1-\rho)x - \alpha\tau(1-x)}{x(1-Ax)} \right]^2$. If $\beta_L = 0$, then $x = \tau$. Substituting $x = \tau$ gives $W^Q = p + \alpha\tau + \frac{2\psi\tau(1-p)(\alpha(1-\tau) - p(1-\rho))^2}{1 - \tau(1 - p + p\rho)}$.

If $\beta_L > 0$, then the low-integrity disingenuous type mixes. By the indifference condition derived in the proof of Proposition 1, $\frac{1}{2} - P(L, L) = c_L$. Since $P(L, L) = \frac{1}{2} + \psi D(H, L)$, we have $D(H, L) = -\frac{c_L}{\psi}$. Therefore, $W^Q = p + \alpha\tau + 2\psi \Gamma_{HL} \left(\frac{c_L}{\psi} \right)^2$. Using $\Gamma_{HL} = (1 - p)x(1 - Ax)$, we obtain $W^Q = p + \alpha\tau + \frac{2c_L^2(1-p)}{\psi} x(1 - (1 - p + p\rho)x)$. This proves the result.

For comparative statics, first consider the case $\beta_L = 0$. Let $S := \alpha(1 - \tau) - p(1 - \rho)$ and $R := 1 - \tau(1 - p + p\rho)$. Then $W^Q = p + \alpha\tau + \frac{2\psi\tau(1-p)S^2}{R}$. Hence

$$\frac{\partial W^Q}{\partial c_L} = 0, \quad \frac{\partial W^Q}{\partial \psi} = \frac{2\tau(1-p)S^2}{R} \geq 0.$$

The other comparative statics are generally conditional:

$$\frac{\partial W^Q}{\partial \alpha} = \tau + \frac{4\psi\tau(1-p)(1-\tau)S}{R},$$

$$\frac{\partial W^Q}{\partial \rho} = \frac{2\psi\tau(1-p)pS(2R + \tau S)}{R^2},$$

and

$$\frac{\partial W^Q}{\partial p} = 1 + \frac{2\psi\tau}{R^2} \left[-S^2R - (1-p)(1-\rho)(2SR + \tau S^2) \right].$$

Thus, when $\beta_L = 0$, welfare is independent of c_L and increasing in ψ , while the effects of α , p , and ρ depend on whether these parameters increase or reduce the utility difference generated by the observed signals.

Now consider the case $\beta_L > 0$. Let $x := \tau + (1-\tau)\beta_L$, $A := 1-p+p\rho$, and $H(x) := x(1-Ax)$. Then $W^Q = p + \alpha\tau + \frac{2c_L^2(1-p)}{\psi}H(x)$. Since x is endogenous, the comparative statics are

$$\frac{\partial W^Q}{\partial \alpha} = \tau + \frac{2c_L^2(1-p)}{\psi}(1-2Ax)\frac{\partial x}{\partial \alpha},$$

$$\frac{\partial W^Q}{\partial \tau} = \alpha + \frac{2c_L^2(1-p)}{\psi}(1-2Ax)\frac{\partial x}{\partial \tau},$$

$$\frac{\partial W^Q}{\partial p} = 1 + \frac{2c_L^2}{\psi} \left[-H(x) + (1-p) \left((1-\rho)x^2 + (1-2Ax)\frac{\partial x}{\partial p} \right) \right],$$

$$\frac{\partial W^Q}{\partial \rho} = \frac{2c_L^2(1-p)}{\psi} \left[-px^2 + (1-2Ax)\frac{\partial x}{\partial \rho} \right],$$

$$\frac{\partial W^Q}{\partial \psi} = -\frac{2c_L^2(1-p)}{\psi^2}H(x) + \frac{2c_L^2(1-p)}{\psi}(1-2Ax)\frac{\partial x}{\partial \psi},$$

and

$$\frac{\partial W^Q}{\partial c_L} = \frac{4c_L(1-p)}{\psi}H(x) + \frac{2c_L^2(1-p)}{\psi}(1-2Ax)\frac{\partial x}{\partial c_L}.$$

By Proposition 4,

$$\frac{\partial x}{\partial \alpha} > 0, \quad \frac{\partial x}{\partial \tau} > 0, \quad \frac{\partial x}{\partial \rho} > 0, \quad \frac{\partial x}{\partial \psi} \geq 0, \quad \frac{\partial x}{\partial p} < 0, \quad \frac{\partial x}{\partial c_L} < 0.$$

Therefore, the comparative statics of W^Q in the mixed equilibrium are generally condi-

tional. The reason is that parameters affect welfare both directly and indirectly through the equilibrium frequency x of vice-signaling.

It remains to verify the effect of c_L in the mixed case. Let $a := \alpha\tau$, $k := p(1-\rho)$ and $A := 1-p+\rho p$, and define $N(x) := a(1-x)-kx$ and $D(x) := x(1-Ax)$. In the mixed equilibrium, $c_L = \psi \frac{N(x)}{D(x)}$. Moreover, $W^Q = p + \alpha\tau + 2\psi(1-p)\frac{N(x)^2}{D(x)}$. Differentiating with respect to x gives $\frac{dW^Q}{dx} = \frac{2\psi(1-p)N(x)}{D(x)^2} [a(2Ax - x - 1) - kx]$. In the mixed equilibrium with $c_L > 0$, we have $N(x) > 0$. Also, $2Ax - x - 1 = x(2A - 1) - 1 < 0$, because $A \leq 1$ and $x < 1$. Therefore $\frac{dW^Q}{dx} < 0$. Since the mixed-equilibrium condition is $B(x) = c_L$ and $B'(x) < 0$, we have $\frac{dx}{dc_L} = \frac{1}{B'(x)} < 0$. Hence $\frac{dW^Q}{dc_L} = \frac{dW^Q}{dx} \frac{dx}{dc_L} > 0$. Thus W^Q is strictly increasing in c_L in the mixed equilibrium whenever $c_L > 0$. $\square$