Moral Values They Believe In: A Model of Electoral Competition*

Tom (Hyeon Seok) Yu[†]

Last Update: June 2019

Abstract

Understanding voter and candidate behavior in elections remains a fundamental question in political economy. This paper develops an electoral competition model with heterogeneity in individuals' party and moral identity. In addition to the formalization of moral values, notable features of the model include (a) the ex-ante correlation between moral and partisan identification and (b) the presence of cheap talkers. The analysis reveals that candidates who can lie have a significant advantage in elections, but the presence of other types of candidates and the voter's endogenous preference for honest candidates constrain the former's pandering behavior. More interestingly, extending the model with the two features produces a similar result, but through different mechanisms, that morally aligned but extremely partisan candidates have a significant chance of winning.

[*] I thank Ken Shotts, Saumitra Jha, Jon Bendor, David Broockman, Steve Callander, Neil Malhotra, Daniel Chen, Zachary Taylor, and Ashutosh Thakur for their helpful guidance and feedback.

[†] Stanford University Graduate School of Business. Email: tom.hs.yu@stanford.edu

"Personally, I think that they need to bring America back to an older value system, where we were about us. We need to come back to a thing where we support Americans. Less abroad and more home."
– A Trump Supporter from Oregon.

"The American story is a journey of continuous striving to more fully realize our founding principles of hope and opportunity for all...Bernie Sanders stands for that America, and so I stand with Bernie Sanders for president."
– A Sanders Supporter from Oregon.

1 Introduction

Why do voters vote the way do, and how do candidates select their campaign strategies? Despite a plethora of research that grapples with these fundamental questions in political economy, they remain as open questions especially with regard to the recent election results in the U.S. and abroad. A class of electoral competition models based on Hotelling (1929) and Downs (1957) has largely focused on policy platforms as the central aspect of candidates that voters consider in elections, but some social scientists have uncovered empirical patterns suggesting that something other than policies drives individuals' voting decisions: values.[1] Indeed, voters' direct accounts of their rationale behind supporting certain candidates hint at the relevance of values in politics, and Enke (2018) provides an empirical evidence that "both voters and politicians exhibit heterogeneity in their emphasis on "universal" relative to "communal" moral values, and that politicians' vote shares partly reflect the extent to which their moral appeal matches the values of the electorate."[2]

Such findings motivate deeper questions on voter and candidate behavior: given that moral values seem to matter in elections, under what conditions do candidates appeal their moral values to voters, and when do voters reward candidates for such an alignment? This paper develops an electoral competition model where individuals have heterogeneous moral types and party identification to address these questions. More specifically, it constructs an electoral setting where candidates choose to campaign on either policy platforms or moral messages, and a representative voter selects a candidate based on the type of campaign message delivered by the former. Information asymmetry arises from the voter's true party

[1] Specific examples include Caprara et al. (2006), Piurko, Schwartz, and Davidov (2011), Schwartz, Caprara, and Vecchione (2010), Sherman (2018), and Enke (2018).

[2] Base on a framework known as Moral Foundations Theory (Haidt 2013), Enke (2018) presents two categories of moral values: "communal" and "universal." Communal values are values that are tied to specific groups or individuals, and examples include loyalty, betrayal, respect, and tradition. Universal values, on the other hand, "apply irrespective of the context or identity of the target person," and examples include equality and fairness.

identification and moral type being common knowledge while candidates' types remain private. To better capture reality, this paper extends the model by allowing individuals' moral preference and party affiliation to be correlated and by introducing cheap talkers who can lie about their types.

This paper makes two main contributions. First, it provides sharper predictions on voter and candidate behavior in a setting where both party and moral alignment can matter. The weight of importance the voter assigns to moral or party alignment generally drives candidates' behavior, but introducing cheap talkers to the model leads the honest candidates to behave in a way that no longer caters to the voter's taste. Interestingly, the presence of such cheap talkers induces the voter to be wary of "too good to be true" announcements, thereby constraining the potential pandering behavior of these liars.[3] In terms of electoral outcomes, cheap talkers are advantaged, although morally-aligned candidates have a significant chance of winning.

Second, the analysis reveals two mechanisms through which extremely partisan but morally-aligned candidates can be elected with a high probability. One mechanism results from the presence of cheap talkers who induce other candidates to behave in ways that consequently leave the voter indifferent between moderate policy platforms and moral messages that show alignment. In the absence of cheap talkers, morally-aligned candidates have a clear preference for campaigning on policies if (a) they are endowed with moderate party identification and (b) the voter cares sufficiently about party alignment. Cheap talkers change this behavior by making any fairly moderate policy announcements equally attractive to the voter, which in turn renders the announcement that signals moral alignment just as attractive. Consequently, any morally-aligned candidates endowed with extreme partisanship can win with at least $\frac{1}{2}$ probability simply by signalling their moral alignment.

Another mechanism results from the extension that correlates individuals' moral types and party affiliation. A key premise behind this correlation is that one's moral identity shapes party identification.[4] The correlation allows the voter to form a stronger (or weaker) belief about candidates based on their campaign messages; if one campaigns on a Republican policy platform, for instance, the candidate is more likely to hold communal moral values and vice versa. Such a learning process induces candidates to campaign less on policies that are

[3] This result aligns with Callander and Wilkie (2007) who also find that voters become "wary of overly-attractive campaign promises, and develop endogenously a preference for more honest candidates."

[4] This premise is based on the moral psychology literature, which largely relies on evolutionary psychology (see e.g., Haidt (2013)).

"less" associated with the voter's moral values. Then, the voter may prefer a morally-aligned but extremely partisan candidate over a candidate with a moderate policy announcement but more likely to be morally-misaligned. This constitutes another mechanism through which extreme candidates can win the election.

The model presented is distinct from existing electoral competition models for several reasons. First, the electoral competition framework of this paper is unique, as it adopts and formalizes a well-established concept of moral identity from moral psychology. Relatedly, operating on the premise that motivates the correlation between types sets the current model apart from other multidimensional voting models that generally assume independence across different characteristics of an individual.[5] The current formulation not only better reflects the reality, but also yields a non-obvious insight on extreme partisans. Lastly, while some results of the paper – the presence of certain types of candidates drastically affecting other candidates' behaviors and policy divergence among candidates – have already been shown in previous research (see e.g., Callander and Wilkie 2007; Kartik and McAfee 2007), explanations for how and why honest extreme candidates can beat moderate ones appear not as readily available. The model's results could help explain an empirical phenomenon on party extremists winning offices (Bonica and Cox 2018).

The remainder of the paper proceeds as follows. Section 2 discusses relevant existing literature. Section 3 presents a baseline model that assumes no correlation between moral and party affiliation and no cheap talkers. After addressing some modelling issues in section 4, sections 5 provides theoretical results including those from extensions that relax the two assumptions. Section 6 discusses the results in depth, and section 7 concludes.

2 Related Literature

This section situates this paper in the political economy literature by considering a number of key existing works in the fields of political economy, political science, and moral psychology. In addition, it introduces empirical regularities that motivate the theory developed.

Electoral Competitions: Liars and Different Types of Valence

Electoral competition has long been a prominent topic in political economy. For the current

[5] Some examples of these multidimensional voting models include Calvert (1985), Kartik and McAfee (2007), Callander and Wilkie (2007), and Bernhardt, Câmara, and Squintani (2011). One potential exception is Callander (2008), whose model ties candidates' political motivations to the effort they exert once entering the office.

endeavor that seeks to better understand the role of moral values in an electoral environment, previous works that consider heterogeneity in types of voters and candidates along with a realistic element of lying by the latter are of particular interest. Such theoretical works are based largely on Banks (1990), which develops a two-candidate electoral competition model where true policy preferences of candidates are private information and candidates can announce positions different from their ideal points by incurring some costs.[6]

Callander and Wilkie (2007) extend Banks (1990)'s model by relaxing the uniform assumption on costs and introducing heterogeneity in both policy intention and the cost candidates incur from "lying" about their policy intention. More specifically, there are high-cost and zero-cost type candidates, and the latter type can lie without incurring any costs. While they find that candidates willing to lie are favored, the mere presence of high-cost (honest) and moderate types preferred by the median voter prevents zero-cost candidates from always pandering to the median voter by pooling at the voter's ideal point. This paper adopts a similar construction by introducing the zero-cost type candidates in an extension.

Kartik and McAfee (2007) present another formal model of electoral competition where candidates' valence – namely, character – in addition to policy preferences matters; those with character suffer "infinite disutility from proposing a platform they do not "believe in," which render them as non-strategic types who always campaign on their true policy preferences. Voters, whose utility depends on both candidate policy platform and character, generally prefer candidates with character and true policy preferences close to theirs. Two aspects of their model are particularly relevant for the current paper. First, their model introduces another dimension that voters might consider in the real world elections; candidates in their model largely try to portray themselves as ones with character, while not appearing too extreme on the policy front. Second, their construction connects character and policy preferences together. While the main dimension of interest in this paper is moral values, the authors' construction is reflected in the baseline model where all candidates are assumed to be the ones with character.

Moral Values in Politics

The relevance of moral values in politics is nothing new for political scientists and moral psychologists. Lakoff (2002), for instance, argues that the difference in worldview that cen-

[6] The key assumption in the model is that "announcing a position different from the true position is costly to the winning candidate, with these costs increasing as the difference between the true policy and the announced policy increases." In such a setting, moderate candidates are generally advantaged relative to extreme ones, as the former incurs less costs from announcing a position closer to that of the median voter.

ters on two opposing models of the family can explain the difference in issue positions and language adopted by liberals and conservatives.[7] How do these two models differ? The strict father model emphasizes "moral strength (the self-control and self-discipline), respect for an obedience to authority, the setting and following of strict guidelines and behavioral norms," where as nurturant parent model assigns highest priorities to "empathy for others and the helping of those who need help." Deason and Gonzales (2012) analyze 2008 presidential party convention acceptance speeches and find more or less confirmatory evidence for the theory: "Democrats referenced more Nurturant Parent themes than Strict Father themes but that Republicans used instantiations from both moral worldviews at similar rates."

Based largely on evolutionary psychology and anthropology, Haidt (2013) provides a positive framework known as Moral Foundations Theory ("MFT") that captures heterogeneity in individuals' moral identity. There are six moral foundations that form such an identity:[8]

Table 1: Moral Foundations Summary

	Evolutionary Basis (i.e., evolved in response to)	Effect (i.e., sensitive to)
Care/Harm	caring for vulnerable children.	signs of suffering and need.
Fairness/Cheating	reaping the rewards of cooperating without getting exploited.	indications for collaboration and reciprocal altruism.
Loyalty/Betrayal	forming and maintaining coalitions.	signs of a team player and a betrayer.
Authority/Subversion	forging beneficial relationships within social hierarchies.	signs of rank or status and conformity to positions.
Sanctity/Degradation	omnivore's dilemma; presence of pathogens and parasites.	new /diverse array of symbolic objects and threats.
Liberty/Oppression	resisting/removing bullies and tyrants.	signs of attempted domination.

His work finds that liberals generally place higher emphasis on care, fairness, and liberty foundations, while conservatives endorse all six, though they are "more willing than liberals to sacrifice Care...in order to achieve their many other moral objectives."[9] This correlational finding suggests an overlap between Lakoff (2002) and Haidt (2013)'s MFT. Despite some obvious differences in theoretical basis, both find that liberals and conservatives show a systematic difference in weights they place on certain moral values. This constitutes a key empirical regularity that forms the basis of a setup where individuals' moral types and political preferences are correlated in an extended version of the model.

Connecting Moral Psychology and Political Economy

Enke (2018) investigates the relevance of moral values in voting behavior, which has not

[7] The author labels these two models as "strict father" and "nurturant parent" which represent the family-based morality of conservatives and liberals, respectively. The general idea is that individuals' views on proper family structure and objective largely shape political preferences.

[8] Enke (2018) considers the original form of MFT that did not include the liberty/oppression foundation. This paper considers all six based on Haidt (2013)'s modification.

[9] Figure 8.2 (p.187) suggests that conservatives are also willing to sacrifice (i.e., place less weight on) the Fairness foundation.

been particularly well explored in political economy. Indeed, most previous theoretical and empirical work in the electoral competition literature have incorporated policy positions and different types of valence that are theoretically broad (e.g., "high" types) and empirically imprecise (e.g., incumbency advantage, education, campaign funding capabilities as proxies for competence).[10]

Despite its novelty, the conceptual framework of his paper makes one omission: party identification of voters and candidates. Do voters make their electoral decisions solely based on moral values?[11] A figure that shows the relative frequency of communal moral terminology among presidential candidates during primaries suggests not.[12] That is, moral alignment seems relevant for the 2016 primaries, but not the previous ones; candidate Obama defeated other democratic candidates in 2008 despite being more communal than universal in his moral language and similarly for candidate McCain and candidate Romney. Albeit only suggestive, one can reason from such a pattern that moral values cannot be deemed as the sole determinant of voting decisions. This paper closes this gap by making both partisan and moral alignment matter.

Although not directly related, this paper also fits into a set of interdisciplinary endeavors that formalize empirical findings and insights from cognitive psychology, sociology, and other disciplines including Besley and Persson (2019), Ortoleva and Snowberg (2015), Enke (2017), and Ryan (2014).

3 Theory & Model

The main objective of the paper is to better understand behaviors of voters and candidates with heterogeneous party and moral identity in an electoral setting. This section opens with the verbal theory that constitutes the basis of a formal electoral competition model subsequently developed.

Verbal Theory

Similarity likely plays a central role in voters' decision making process. The key observation underlying this premise is that humans exhibit homophily. Sociologists have found that

[10] More specific examples include Groseclose (2001), Ansolabehere and Snyder (2002), Meirowitz (2007), Adams et al. (2011), Stone and Simas (2010), and Bernhardt, Câmara, and Squintani (2011).

[11] Enke (2018)'s voter utility function does include "economic incentive" term, but voters could presumably have policy preferences that do not align with their economic incentives.

[12] See Figure 3 at p.17.

"similarity breeds connection" and denoted this as the homophily principle (see e.g., Fu et al. 2012; McPherson, Smith-Lovin, and Cook 2001). In particular, individuals have a "tendency...to associate with those of their own political orientations."[13] In addition to driving personal connections, homophily seems to affect individuals' political decisions/perceptions: Bailenson et al. (2008), for instance, find that voters prefer facially similar candidates when they are not familiar with such candidates in an experimental setting. Presumably, the type of similarity that matters in elections (especially those with a low media coverage) might be party identity; voters care whether a given candidate represents the same party and prefer to elect the one with the same party affiliation, ceteris paribus.[14]

Then, where do moral values become relevant in voters' decision making process, if at all? Psychologists have long noted that individuals experience psychological distress from acknowledging/thinking that they have done something wrong; they like to believe that they are correct or are doing the "right" thing.[15] Moral values, which have evolutionary bases, serve as internal evaluation standards that help individuals distinguish right from wrong.[16] Assuming that voters have the objective of selecting the right candidate so as to minimize the potential dissonance resulting from choosing a wrong one, then, voters might consider moral values for which candidates appear to stand as another evaluation criteria in their decision making process in addition to candidates' party affiliation. Accordingly, candidates would signal to voters that they share similar values and party identity.

In sum, the central premise based on the empirical regularity of homophily is as follows: voters prefer, hence select, those who are similar to them. More specifically, this serves as a basis for why voters in the formal model of this paper prefer candidates whose moral values and party identity align with theirs.

Formalization – Model Setup

[13] The relevant citations are noted in McPherson, Smith-Lovin, and Cook (2001). More recently, Huber and Malhotra (2016) find that individuals' partner preferences are also driven by political preferences based on their analysis of individuals' online dating behavior.

[14] It is notable that canonical electoral competition models including Hotelling (1929) and Downs (1957) also reflect this notion of homophily, as voters tend to prefer candidates whose policy preferences are closer to their own.

[15] More technically put, this is to minimize cognitive dissonance. For the canonical discussion of cognitive dissonance, see Festinger (1962).

[16] According to evolutionary anthropologists including Henrich (2015), some moral values such as in-group loyalty might have developed based on what they call "norm psychology" that leads us to (1) "at a young age... develop cognitive abilities and motivations for spotting norm violations and avoiding or exploiting norm violators, as well as for monitoring and maintaining our own reputations" and (2) "internalize" norms as goals in themselves.

There are three main actors: one representative voter and two candidates, each initially endowed with party and moral identities. Denote an individual i's true party and moral types as p_i and t_i, respectively. The party identification space is continuous; $P \in [-1, 1]$, and assume the midpoint of P is zero. The moral space, on the other hand, is discrete; $t_i \in \{O, U\}$ where O denotes communal type and U, universal. The prior probability of an individual being either type is symmetric (i.e., $Pr(t_i = O) = Pr(t_i = U) = \frac{1}{2}$).

Similar to Banks (1990) and Callander and Wilkie (2007), individuals in this model have party identification that are assumed to be independent and identically distributed random variables with cdf $F(\cdot)$ and density $f(\cdot)$, which is uniform and symmetric about zero. In the context of the US politics, for example, an individual i endowed with $p_i = 1(-1)$ represents that one is an extreme Republican (Democrat); this could be interpreted as having a policy preference that agrees with the averaged party members' preferences on every issue.[17] Candidates' party and moral types are private information, whereas the representative voter's type is assumed to be common knowledge. Such asymmetry is based on descriptive evidence that candidates generally conduct thorough research on the electorate, while most voters do not have time nor willingness to study candidates (see e.g., Annenberg Public Policy Center 2014, Somin 2014). However, the party identity distribution $f(\cdot)$ and the prior probability on moral types are common knowledge. Also note that candidates themselves do not know their opponents' true types, since they are assumed to be private information.

The sequence of the game is as follows:

1. Nature selects the representative voter's and candidates' party affiliation and moral types.
2. Both candidates observe the voter's type. They select campaign message contents as either policy platforms (partisan messages) or moral appeals.
3. The voter observes candidates' campaign messages and then casts her vote.
4. Winner is selected and enacts policies.[18]

Again, the central premise of the model is that voters prefer, hence select, those who are similar to them. Accordingly, a voter i's utility from selecting candidate j is defined as

[17] Put another way, the current formulation of the party identity represents how partisan (i.e., how loyal to a given party) a given individual is.

[18] This necessarily means that the partisan component of the voter's utility is realized, while the same need not necessarily hold for the moral component. That is, the voter may end up never learning the winner's true moral type.

follows:

$$u_i(p_j, t_j) = -\lambda(p_j - p_i)^2 - (1 - \lambda)\mathbb{1}(t_j \neq t_i)$$

As suggested in the sequence, the main action taken by the candidates is the selection of campaign message content, which is assumed to be binary; denote policy (partisan) and moral messages delivered by candidate j as $(x_j, m_j) \in \{(\varnothing, t_j), (p_j, \varnothing)\}$. That is, if a candidate selects to campaign on policy or partisan message, then he can only send policy (partisan) message, and the voter does not directly observe his moral type. Define $\gamma(p_i, t_i) \in \{0, 1\}$ as the binary strategy function that represents candidates' choice of the message content given voter i's party and moral types; $\gamma_j = 1$ denotes candidate j campaigning entirely on policy (partisan) platforms, and $\gamma_j = 0$ denotes sending moral messages. Candidate j's utility from winning the election is as follows:

$$u_j = W - c$$

where W is the office benefit, and c is the cost of participating in the election.[19] This means candidates are all "office-motivated," as they derive utility from winning rather than implementing their preferred policies in the office.

A voter's ex-ante expected utility from a candidate j whose party and moral types are not known (i.e., before seeing their messages) is:

$$E[u_i|(x_j, m_j) = (\varnothing, \varnothing)] = -\lambda(E[p_j] - p_i)^2 - \sigma^2 - \frac{1 - \lambda}{2}$$

where $\sigma^2_{p_j}$ is the variance associated with the density of voters' beliefs regarding candidate j's true party identification, and these beliefs are assumed to be common across voters. Given that voters cast votes upon seeing candidates' messages, define $r_i((x_1, m_1), (x_2, m_2))$ as the strategy function that represents the probability of voter i selecting candidate 1 upon observing campaign messages from both candidates. Accordingly, the probability that i votes for candidate 2 is $1 - r_i((x_1, m_1), (x_2, m_2))$. Voters use weakly undominated strategies, meaning they select a candidate they strictly prefer and toss a fair coin when indifferent.

Finally, denote voters' posterior beliefs on candidate j's party identity distribution upon seeing a moral message as $\phi[(x_j, m_j)]$. Similarly, define voters' posterior beliefs on candidate j's moral type being misaligned with that of the voter i upon seeing a policy (partisan) platform p_j as $\mu[t_j \neq t_i|(x_j, m_j) = (p_j, \varnothing)]$.

[19] As the question at hand pertains to the candidate behavior rather than entry, I assume that $W > 0$ and that W is sufficiently greater than c to ensure that the two candidates always choose to run.

4 Modeling Issues

Before delving into the analysis, I discuss three aspects of the current setup some may find odd or troubling.

Party Identity instead of Ideological Identity

One distinct feature of the current model is that individuals are endowed with a party identity (e.g., Democrat-Republican on a continuous spectrum in the US context) rather than a conservative-liberal ideological preference. This was a deliberate modeling decision based on numerous empirical evidence that many voters are ideologically-mixed – they lack ideological consistency across different issues (Broockman 2016; Baldassarri and Gelman 2008; Converse 1964; Kinder and Sears 1985). Kinder and Kalmoe (2017), in particular, find that voters' indication of ideological identity changes rather frequently in a short period of time (two years), whereas their party identification showed a much greater degree of stability.

Given such patterns, a conventional approach that assumes an individual to be endowed with a constant uni-dimensional ideological preference does not seem to best reflect the reality. Therefore, the current model assumes an individual to be endowed with a party identity instead; the real number from the interval $[-1, 1]$ represents how strong of a partisan a given individual is. For candidates, then, this means their campaign messages signal their degree of partisanship and moral values, not how liberal or conservative they are.

Restricted Moral Types

Most moral psychologists would agree that an individual's moral identity is neither binary nor mutually exclusive. Indeed, with regard to the six moral foundations under MFT, an individual can strongly agree (disagree) with all six, which would necessarily mean that an individual is both (neither) communal and (nor) universal.

Yet, the current construction restricts actors' moral identity to be either communal or universal, and it does not allow one to be both; being a communal type necessarily means one is not universal. I justify this rather strong assumption with tractability and, to a weaker extent, reality. First, restricting the type simplifies the formal analysis and allows derivation of interpretable equilibria, a desirable benefit for parsing out the fundamental electoral mechanisms that drive behaviors.[20] Second, survey data on individuals' endorsement of the five

[20] A similar justification applies to the modelling decision of restricting candidates' actions to be binary (i.e., either entirely policy platforms or moral messages).

foundations shows a relatively clear pattern that individuals who show strong endorsement for care and fairness (universal values) show comparatively weaker endorsement for loyalty, authority, and sanctity (communal values), and the opposite holds for those who strongly endorse the latter (Haidt 2013). Such a pattern renders the current model not entirely restrictive or unreasonable.

Lack of Intermediate Actors (e.g., Media)

Iyengar (2005) casts doubt on the explanatory power of Lakoff (2002)'s verbal theory by pointing out the lack of consideration for "how political issues are framed" and become salient by intermediary actors like media in elections. His criticism certainly applies to the current model, as media or any other actors who can affect the voter's perception of candidates aside from candidates themselves are not modeled at all. For the current nontrivial objective of understanding candidate and voter behavior in the world of heterogeneous policy and moral preferences, however, such an extension remains out of scope. Nevertheless, the critique is valid, and the extension based on the current framework should be feasible.[21]

5 Theoretical Predictions

5.1 Baseline Results

This section provides results from solving the model with the following strong assumption that restricts candidate behavior: **candidates are assumed to suffer infinite costs from lying.**[22] Denote the representative voter v, whose moral type is $t_v = O$ and policy preference $p_v = 0$. For candidates, label the morally-aligned candidates (i.e., those who share the same moral values) as candidate type A and misaligned, hence disadvantaged, as candidate type D. In this setting, type A candidates' true type is a tuple (p_A, O) and type D candidates (p_D, U). To conserve notation, the voter strategy r_v denotes the probability that the voter selects the first candidate she sees (i.e., candidate 1). The results are organized by the three cases of the voter's weight on policy versus moral alignment (λ). All proofs are provided in the Appendix.

Case 1: the voter only cares about moral alignment ($\lambda = 0$)

This is one of the corner cases where the voter only considers moral alignment of the can-

[21] More concretely, these intermediary actors could be introduced by allowing them to directly affect the magnitude of λ; the degree to which voter cares about moral or certain policy issues in an election.

[22] From Kartik and McAfee (2007)'s formulation, this is equivalent to assuming all candidates to have "character."

didate. Perhaps unsurprisingly, such a taste induces A types to always campaign on moral messages, while leaving D types indifferent between either action. The intuition behind this result is straightforward: D types, who cannot lie while being aware of the moral mismatch, are indifferent between either action, since the voter can infer that their values are not communal from any of their actions. For A types, capitalizing on their moral advantage constitutes a best response in this setting.

***Result 1**: Suppose the voter only cares about moral alignment ($\lambda = 0$). An equilibrium can be characterized as follows:*

- $\gamma_A^* = 0 : (x_A^*, m_A^*) = (\varnothing, O)\ \forall p_A \in P$; *morally advantaged candidates (A) always campaign on moral messages.*
- $\gamma_D^* = \begin{cases} 1 : (x_D^*, m_D^*) = (p_D, \varnothing), & \textit{with probability } \frac{1}{2} \\ 0 : (x_D^*, m_D^*) = (\varnothing, U), & \textit{with probability } \frac{1}{2} \end{cases} \forall p_D \in P$; *morally disadvantaged candidates (D) are indifferent between campaigning on policy or moral messages.*[23]
- $r_v^* = \begin{cases} 1, & \textit{if } (x_1, m_1) = (\varnothing, O) \neq (x_2, m_2) \\ \frac{1}{2}, & \textit{if } \begin{cases} (x_1, m_1) = (x_2, m_2) \\ (x_1, m_1) \neq (x_2, m_2) \wedge m_1 = m_2 = \{\varnothing\} \end{cases} \\ 0, & \textit{otherwise.} \end{cases}$
- $\mu_v^*[t_j \neq t_v | (x_j, m_j) = (p_j, \varnothing)] = 1$; *the voter believes candidate j to be morally misaligned if she observes policy platforms instead of moral messages.*

A types have a strictly greater probability of winning the election, while D types' chance of winning is a coin-toss if the opponent also holds universal values. The subsequent extensions that correlate types and introduce zero-cost liars show that this result remains largely unchanged.

Case 2: the voter only cares about policy alignment ($\lambda = 1$)

Note that neither type of candidate has an advantage in this case and that not revealing one's policy preference still conveys information about one's policy type because the voter believes that a candidate revealing his moral side is likely to have an extreme policy preference. This means that a cut-point equilibrium where some candidates with policy preferences below some threshold sending policy messages while others sending moral messages cannot be sustained, as there could be "moderately extreme" policy types over a certain threshold

[23] Strictly speaking, no strategy is weakly dominated for D types; they are expected to lose to morally-aligned candidates regardless of their actions. Since the voter does not care about policy at all, the "goodness" of one's policy type does not improve the relative appeal of either action.

who can profitably deviate by sending policy message instead. As a result, all candidates' best response is to campaign on policy platforms, and the voter's posterior that assumes the worst policy type for the off-path equilibrium behavior (i.e., campaigning on moral messages) can sustain this equilibrium.

***Result 2**: Suppose the voter only cares about policy alignment ($\lambda = 1$). A unique equilibrium exists in which:*

- $\gamma_A^* = \gamma_D^* = 1 : (x_A^*, m_A^*) = (p_A, \varnothing) \wedge (x_D^*, m_D^*) = (p_D, \varnothing)\ \forall p_A, p_D \in P$; *both types of candidates always campaign on policy.*
- $r_v^* = \begin{cases} 1 & \textit{if} \begin{cases} (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (p_2, \varnothing) \wedge |x_1| < |x_2| \\ (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (\varnothing, t_2) \end{cases} \\ \frac{1}{2}, & \textit{if } (x_1, m_1) = (x_2, m_2) \\ 0, & \textit{otherwise.} \end{cases}$
- $\phi_v^*[(x_j, m_j) = (\varnothing, t_j)] = 1$; *the voter's posterior on candidate j's policy type upon observing an off-the-path behavior considers the candidate's policy platform to be the worst possible type.*

Just as in *Result 1*, the voter's taste largely drives candidates' behavior; they campaign only on policy platforms. With the moral advantage being irrelevant, the candidate with a policy preference closer to that of the voter wins the election.

Case 3: the voter cares about both ($\lambda \in (0, 1)$)

This part considers a much more realistic case where both policy and moral alignment matter, and the formal analysis reveals multiple equilibria. This part focuses on two of the four equilibria identified, as candidates' behavior in these two equilibria differs from those of the two cases above. First, there is a separating equilibrium where all A types campaign on moral messages, while all D types campaign on policy platforms. Simply put, the advantaged types find it a best response to utilize their moral alignment, and this requires the voter to care sufficiently about moral alignment to sustain this equilibrium (i.e., an upper bound on λ).

In another equilibrium, even some A types – namely, those endowed with fairly moderate policy preferences – campaign on policy platforms, while D types continue to campaign on policies. For a symmetric reason as the separating equilibrium, λ should be sufficiently high to sustain such an equilibrium.

Result 3: *Suppose the voter cares about both policy and moral alignment ($\lambda \in (0,1)$). There exists multiple equilibria:*

1. ***Separating***. *Suppose $\lambda \leq \frac{1}{1+\sigma^2} = \frac{3}{4}$. The following constitutes an equilibrium:*

   - $\gamma_A^* = 0 \ \forall p_A \in P$; *A types always campaign on moral messages.*
   - $\gamma_D^* = 1 \ \forall p_D \in P$; *D types always campaign on policy platforms.*
   - $r_v^* = \begin{cases} 1 \quad \text{if} \begin{cases} (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (p_2, \varnothing) \wedge |x_1| < |x_2| \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (p_2, \varnothing) \end{cases} \\ \frac{1}{2}, \quad \text{if } (x_1, m_1) = (x_2, m_2) \\ 0, \quad \text{otherwise.} \end{cases}$
   - *Voter posterior beliefs:*
     - $\phi_v^*[(\varnothing, U)] = 1$; *the voter's posterior on a morally-misaligned candidate's off-path behavior.*
     - $\phi_v^*[(\varnothing, O)] = U[-1,1]$; *the voter's posterior on morally-aligned type's policy preference based on Bayes's rule.*
     - $\mu_v^*[t_j \neq t_v | (x_j, m_j) = (p_j, \varnothing)] = 1$; *the voter believes candidate j to be morally misaligned upon observing that the candidate campaigns on policy platforms.*

2. ***Semi-Pooling***. *Suppose $\lambda \geq \frac{4}{3D^2+4} = \frac{4}{7}$. The following constitutes an equilibrium:*

   - $\gamma_A^* = \begin{cases} 1, \quad \text{if } |p_A| \in [0, \bar{p}]; \bar{p} \equiv \frac{1}{4}(1 + \frac{\sqrt{3}\sqrt{7\lambda - 4}}{\sqrt{\lambda}}) \\ 0, \quad \text{otherwise.} \end{cases}$; *A types' actions depend on their policy preference realizations.*
   - $\gamma_D^* = 1 \ \forall p_D \in P$; *D types always campaign on policy platforms.*
   - $r_v^* = \begin{cases} 1 \quad \text{if} \begin{cases} (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (p_2, \varnothing) \wedge |x_1| < |x_2| \\ (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (\varnothing, O) \wedge |x_1| < \bar{p} \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (\varnothing, U) \end{cases} \\ \frac{1}{2}, \quad \text{if} \begin{cases} (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (\varnothing, O) \wedge |x_1| = \bar{p} \\ (x_1, m_1) = (x_2, m_2) \end{cases} \\ 0, \quad \text{otherwise.} \end{cases}$
   - *Voter posterior beliefs:*
     - $\phi_v^*[(\varnothing, U)] = 1$; *the voter believes candidate j's policy type to be the worst possible type upon observing moral misalignment.*
     - $\phi_v^*[(\varnothing, O)] = U[-1, -\bar{p}] + U[\bar{p}, 1]$; *the voter's posterior on A types' policy preference is a truncated uniform distribution.*
     - $\mu_v^*[t_j \neq t_v | x_j \in [0, \bar{p}]] = \frac{1}{2}$; *the voter's posterior on candidate j's moral type upon observing that the candidate campaigns on policy platforms, which fall in the interval $[0, \bar{p}]$.*

   - $\mu_v^*[t_j \neq t_v | x_j \in (\bar{p}, 1]] = 1$; *the voter believes candidate j to be morally misaligned upon observing that the candidate campaigns on policy platforms that fall in the interval $(\bar{p}, 1]$.*

3. *If $\lambda \in (\frac{4}{7}, \frac{3}{4})$, then both equilibria can exist.*

In the separating equilibrium, A types are certainly advantaged, as the voter places a sufficiently high weight on moral alignment. Therefore, their best response is to appeal their moral alignment to the voter. D types' best response, unlike *Result 1*, is to always announce their policy preferences due to the voter's off-path belief that a candidate has the most extreme policy preference if he ever announces his (misaligned) moral type.

The semi-pooling equilibrium can exist when the voter places a sufficiently large weight on policy alignment. In this case, some A types whose policy types are sufficiently close to that of the voter consider it a best response to announce their policy positions even at the cost of not being distinguishable from D types. In a situation where both candidates campaign on policies, the voter selects the candidate with a policy platform announcement closer to her ideal point. Furthermore, D types with sufficiently moderate policy types can even beat A types with extreme policy preferences, as the former can take advantage of the fact that the voter cannot be sure of their moral misalignment anymore with some A types also campaigning on policy platforms.

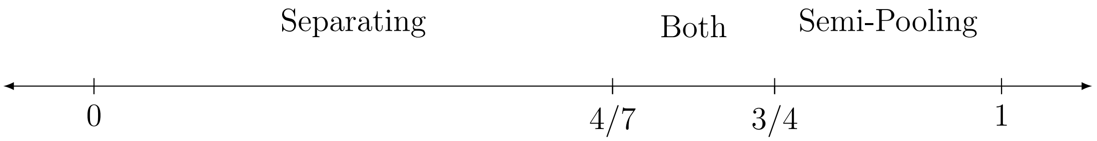


Figure 1: Equilibrium by λ

Boundary conditions on the magnitude of λ suggest that with fairly standard distributions for policy preferences, there exists an overlapping area where both equilibria can exist.[24] The following proposition compares the voter's expected welfare under the two equilibria.

Proposition 1: *When $\lambda \in (\frac{4}{7}, \frac{3}{4})$, the voter's expected utility under the separating equilibrium is strictly greater than that under the semi-pooling equilibrium. Formally,*

$$E[u_v|SE] = \frac{3}{4}[-\lambda\sigma^2] = -\frac{\lambda}{4} > -\frac{\lambda}{4}\{(\bar{p})^3 - \bar{p} + 1\} - (1-\lambda)\frac{\bar{p}}{2}(1 - \frac{(\bar{p})^2}{2}) = E[u_v|SPE]$$

[24] Appendix section C considers the cases of a general uniform distribution and the normal distribution.

where SE and SPE refer to the separating and semi-pooling equilibrium, respectively, and the inequality simplifies to

$$(1-\lambda)(1-\frac{\bar{p}}{2}) > \frac{\lambda}{2}(1-(\bar{p})^2)$$

The inequality holds for all $\lambda \in (\frac{4}{7}, \frac{3}{4})$.[25] This result may appear counterintuitive, as one might expect the voter's utility to be higher when the A type, in a sense, accommodates the voter's care for policy alignment by campaigning on policy instead of moral values. The LHS of the inequality, which represents the expected loss from potentially electing a D type, sheds light on the result; the expected gains from inducing even A types to campaign on policy platforms (i.e., RHS) are not large enough to offset the loss from potentially electing D types.

5.2 Extension 1: Correlated Types

The baseline construction assumes an individual's policy and moral types to be independent, which is inconsistent with previous empirical findings. Indeed, it does not reflect the generally observed regularity that liberals and conservatives exhibit a systematic difference in weights they place on certain moral values (Enke 2018; Haidt 2013; Graham, Haidt, and Nosek 2009). This section extends the model to better reflect this reality by conditioning an individual's policy preferences based on moral types. Formally,

$$Pr(p_i \in \text{U}[-1,0]|t_i = U) = Pr(p_i \in \text{U}[0,1]|t_i = O) = \pi \in (\frac{1}{2}, 1]$$

This conditional probability reflects an assumption that **an individual's moral identity shapes his/her political preferences.** More specifically, being a communal moral type makes one more likely to be on the right side of the political spectrum than the other, and conversely for the universal moral types. Analyzing this modified model shows that results from the baseline model generally hold.

Cases 1 and 2 ($\lambda = 0$ and $\lambda = 1$)

The same equilibrium characterizations from the previous section apply for the two corner cases where the voter entirely cares about either moral or policy alignment; correlating types does not affect the outcome when voter only considers one dimension.

Case 3 ($\lambda \in (0,1)$)

For the similar reason as above, the equilibrium characterization for the separating equilib-

[25] More precisely, the condition holds for all $\lambda \in (\frac{4}{7}, 1)$.

rium remains unchanged. When the voter expects candidates to pool by moral types (i.e., A types campaigning on moral messages and D types on policy platforms), the additional information based on the type correlation becomes irrelevant. Note that the same upper bound on λ applies; the voter must care sufficiently about moral alignment to sustain this equilibrium.

Allowing types to be correlated changes the semi-pooling equilibrium in an intuitive yet important manner. Specifically, the cut-point at which A types are willing to campaign on policy becomes asymmetric. That is, A types who are endowed with some policy $p_A \in [0, 1]$ willing to campaign on policy platforms might not be willing do so if they were instead endowed with $-p_A$, the equidistant policy on the other side.

Result 4*: Suppose the voter cares about both policy and moral alignment ($\lambda \in (0, 1)$). The semi-pooling equilibrium can be characterized as follows:*[26]

- $\gamma_A^* = \begin{cases} 1, & \textit{if } |p_A| \in [\bar{p}_1, \bar{p}_2], |\bar{p}_1| < |\bar{p}_2| \\ 0, & \textit{otherwise.} \end{cases}$*; morally-advantaged candidates' actions depend on their policy preference realizations.*

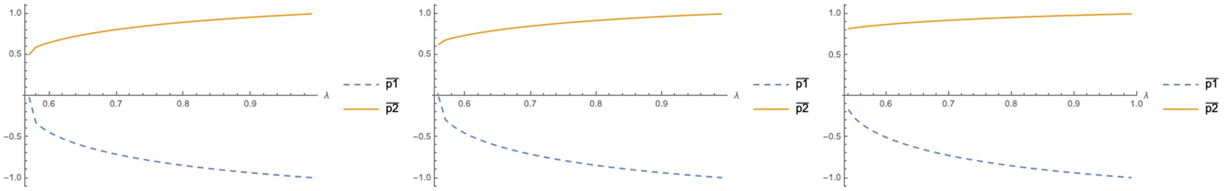
- $\gamma_D^* = 1 \ \forall p_D \in P$*; morally-disadvantaged candidates always campaign on policy platforms.*

- $r_v^* = \begin{cases} 1 & \textit{if } \begin{cases} (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (p_2, \varnothing) \wedge |x_1| < |x_2| \\ (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (\varnothing, O) \wedge x_1 \in (\bar{p}_1, \bar{p}_2) \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (\varnothing, U) \end{cases} \\ \frac{1}{2}, & \textit{if } \begin{cases} (x_1, m_1) = (p_1, \varnothing), (x_2, m_2) = (\varnothing, O) \wedge x_1 = \bar{p}_1 \textit{ or } \bar{p}_2 \\ (x_1, m_1) = (x_2, m_2) \end{cases} \\ 0, & \textit{otherwise.} \end{cases}$

- *Voter posterior beliefs:*
  - $\phi_v^*[(\varnothing, U)] = 1$*; the voter believes candidate j's policy type to be the worst possible type upon observing moral misalignment.*
  - $\phi_v^*[(\varnothing, O)] = (1 - \pi)U[-1, -\bar{p}_1] + \pi U[\bar{p}_2, 1]$*; the voter's posterior on an A type's policy preference is a truncated uniform distribution.*
  - $\mu_v^*[t_j \neq t_v | x_j \in (0, \bar{p}_2]] = 1 - \pi$*; the voter's posterior on candidate j's moral type upon observing that the candidate campaigns on policy platforms, which fall in the interval* $(0, \bar{p}_2]$.

[26] The analytical solutions for cut-points $\bar{p}_1$ and $\bar{p}_2$ are omitted to conserve space. Instead, the figures below provide numerical solutions.

- $\mu_v^*[t_j \neq t_v | x_j \in [\bar{p}_1, 0)] = \pi$*; the voter's posterior on candidate j's moral type upon observing that the candidate campaigns on policy platforms, which fall in the interval* $[\bar{p}_1, 0)$.
- $\mu_v^*[t_j \neq t_v | (0, \varnothing)] = \frac{1}{2}$*; the voter's posterior on candidate j's moral type upon observing that the candidate campaigns on policy platform* 0.
- $\mu_v^*[t_j \neq t_v | x_j \in [-1, \bar{p}_1) \vee (\bar{p}_2, 1]] = 1$*; the voter believes candidate j to be morally misaligned upon observing that the candidate campaigns on policy platforms that fall in the intervals* $[-1, \bar{p}_1)$ *or* $(\bar{p}_2, 1]$.



(a) $\pi = \frac{2}{3}$ (b) $\pi = \frac{3}{4}$ (c) $\pi = \frac{7}{8}$

Figure 2: Numerical Solutions of $\bar{p}_1$ and $\bar{p}_2$ for Various π

Numerical solutions in figure 2 show an intuitive result of $|\bar{p}_1| < |\bar{p}_2|$, while both cut-points approach 1 as λ goes to 1. The correlation in types can induce candidates with aligned moral values but "opposite" policy preferences to campaign on moral messages despite relatively moderate policy preferences. Another notable pattern in the figure is that the range of λ for which an equilibrium can be sustained increases in π. In addition, the cut-point $\bar{p}_2$ is also increasing in π, which is intuitive; higher the correlation between types, more likely it is for a candidate to be of the associated moral type, so candidates find it sufficient to campaign on policy platforms. Finally, the lower bounds on λ that are significantly less than $\frac{3}{4}$ show that there exists an interval where both separating and semi-pooling equilibria can exist, just as in the baseline construction.

As for the outcome of the election, A types continue to have an advantage from their moral alignment, but just as in *Result 3*, D types with policy preferences close to that of the voter can win the election.

5.3 Extension 2: Introducing Cheap Talkers (Zero-Cost Liars)

The baseline construction had a rather strong assumption that all candidates incur infinite costs from lying. Based on Callander and Wilkie (2007) and Kartik and McAfee (2007), this section introduces "zero-cost" types – those who do not incur any costs from lying about their true types. Just as any other candidates, however, such zero-cost type candidates are endowed with some policy and moral types from the same distributions, and the same restriction on actions applies; they can only campaign on either policy or moral messages. Denote such types of candidates as Z, and the probability of being a zero-cost candidate is $Pr(Z = 1) = q \in (0, 1)$. Also, denote the honest types as I for incurring infinite costs. Then, $\gamma_{A,Z}$ represents the strategy of a candidate whose moral type aligns with that of the voter while being a zero-cost type. Finally, define $\eta[Z|(x_j, m_j)]$ as the voter's posterior on the probability of a candidate being a zero-cost type upon observing his campaign messages. Note that this section does not assume correlated types from extension 1.

Case 1: $\lambda = 0$
The general result in this special case remains essentially the same as *Result 1*; with the voter only considering moral alignment, Z types' best response is to pretend to be aligned. The same logic from the previous construction applies for the honest (i.e., infinite-cost type) candidates.[27]

Case 2: $\lambda = 1$
While the introduction of Z types does not affect the general behavior of candidates (i.e., all of them campaign on policy platforms), the equilibrium behavior of Z types is notable. This result is very similar (if not the same) to the unique mixed strategy equilibrium from Kartik and McAfee (2007): given the continuous policy type space and the voter caring only about policy alignment, Z types cannot pool at any particular point; if they do, one can profitably deviate by campaigning on a platform just a bit away from such a point and be perceived as an honest type by the voter. This leaves with them "mixing" over an interval close enough to the voter's ideal point so that the voter prefers to select such close candidates even when knowing that one might elect a cheap talker. Denote the cut-point of this interval $\hat{p}$. The formal characterization of the equilibrium is as follows.

Result 5: *Suppose the voter only cares about policy alignment ($\lambda = 1$). A mixed-strategy equilibrium exists in which:*

- $\gamma^*_{A,I} = \gamma^*_{D,I} = 1 : (x^*_{A,I}, m^*_{A,I}) = (p_{A,I}, \varnothing) \wedge (x^*_{D,I}, m^*_{D,I}) = (p_{D,I}, \varnothing)\ \forall p_{A,I}, p_{D,I} \in P$; *both A, I and D, I types always campaign on policy.*
- $\gamma^*_{A,Z} = \gamma^*_{D,Z} = 1 : (x^*_{A,I}, m^*_{A,Z}) = (x^*_{D,Z}, m^*_{D,Z}) = (p'_Z, \varnothing)\ \forall p_{A,Z}, p_{D,Z} \in P$, *where* $|p'_Z| \in [0, \hat{p}]$ *with mixing probability* $f(p) \equiv \frac{(1-q)(B-p^2)}{q(\frac{1}{3}-B)}$, *where* $B \equiv (\hat{p})^2$ *is the constant expected utility that the voter receives from any policy announced in the given interval.*

[27] Proof and the characterization of the equilibrium is provided in the Appendix.

$\hat{p} \in [0, \frac{1}{\sqrt{3}}]$ *and solves* $q \equiv \frac{4(\hat{p})^3}{4(\hat{p})^3 - 3(\hat{p})^2 + 1}$; *Z types always campaign on policy while mixing over the interval* $[-\hat{p}, \hat{p}]$.

- $$r_v^* = \begin{cases} 1 & \textit{if} \begin{cases} (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| < |x_2| \wedge |x_1| \in [0, \hat{p}], |x_2| > \hat{p} \\ (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| < |x_2| \wedge |x_1|, |x_2| > \hat{p} \\ (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (\varnothing, t_2) \end{cases} \\ \frac{1}{2}, & \textit{if} \begin{cases} (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1|, |x_2| \in [0, \hat{p}] \\ (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| = |x_2| \\ (x_1, m_1) = (x_2, m_2) = (\varnothing, t) \end{cases} \\ 0, & \textit{otherwise.} \end{cases}$$

- *Voter posteriors:*
  - $\phi_v^*[(x_j, m_j) = (\varnothing, t_j)] = 1$; *the voter believes candidate j's policy type to be the worst possible type upon observing a moral message.*
  - $\eta_v^*[Z_j = 1 | (x_j, m_j) = (x_j, \varnothing), x_j \in [0, \hat{p}]] = \frac{qf(x_j)}{qf(x_j) + (1-q)(\frac{1}{2})}$; *the voter's posterior on candidate j being a Z type upon observing a policy platform announcement in the interval* $[0, \hat{p}]$.

Note that for Z types to be indifferent over the interval $[-\hat{p}, \hat{p}]$, it must be that the voter earns the same exact expected utility from any policy platform in the given interval, which in turn implies that Z types are expected to win with a weakly higher probability against I types. The behavior I types remains the same, as their best response in this situation is to campaign on their true policy preferences.

Case 3: $\lambda \in (0, 1)$[28]
Just as in case 2, extending the model to include zero-cost types does not affect the general existence of equilibria from the baseline construction, but the A, I types' behavior in the semi-pooling equilibrium changes drastically. First, the separating equilibrium remains nearly identical to its counterpart in *Result 3.* With the voter sufficiently caring about moral alignment, A, I types find it a best response to take advantage of their moral alignment, and Z types mimic them. Perhaps unsurprisingly, the upper-bound on λ is decreasing in q; as the probability of a candidate being a cheap talker increases, the voter needs to be of a type who places a higher weight on moral alignment to tolerate such an equilibrium.

The semi-pooling equilibrium has a very similar characteristic as *Result 5* in terms of the zero-cost types' behavior. But with the voter also caring about moral alignment, now they

[28] Just as in the baseline model, there are two other equilibria that share the same characterizations as cases 1 and 2. Proofs of these equilibria are provided in the Appendix.

mix at two levels: first, they are conjectured to be indifferent between sending the aligned moral message and "good" policy platforms (i.e., within the boundary of the voter's ideal point). Second, if sending the latter, they need to be mixing over the possible platforms within the interval for the same reason of preventing profitable deviations. In other words, the voter's expected utility from either action should be the same.

Interestingly, such a strategy implies that A, I types endowed with "good" policy preferences (i.e., within the interval $[0, \hat{p}]$ in absolute terms) should also be indifferent between either action. In the previous construction, the voter's higher weight on policy alignment induced even the morally-advantaged type to, in a sense, accommodate such a preference by campaigning on policies. With the introduction of liars, however, the voter can no longer trust the policy platforms shared by candidates, consequently preventing A, I types from campaigning on policies with probability 1.

Result 6: *Suppose the voter cares about both policy and moral alignment ($\lambda \in (0,1)$). There remains multiple equilibria:*

1. ***Separating***. *Suppose $\lambda < \frac{3}{1+q+3}$. The following constitutes an equilibrium:*
   - $\gamma^*_{A,I} = 0\ \forall p_{A,I} \in P$; *A types always campaign on moral messages.*
   - $\gamma^*_{D,I} = 1\ \forall p_{D,I} \in P$; *D types candidates always campaign on policy platforms.*
   - $\gamma^*_{A,Z} = \gamma^*_{D,Z} = 0\ \forall p_{A,Z}, p_{D,Z} \in P$; *Z types always campaign on moral messages.*
   - $r^*_v = \begin{cases} 1 & \textit{if} \begin{cases} (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| < |x_2| \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (x_2, \varnothing) \end{cases} \\ \frac{1}{2}, & \textit{if } (x_1, m_1) = (x_2, m_2) \\ 0, & \textit{otherwise.} \end{cases}$
   - *Voter posterior beliefs:*
     - $\phi^*_v[(\varnothing, U)] = 1$; *the voter believes candidate j's policy type to be the worst possible type upon observing a moral message.*
     - $\phi^*_v[(\varnothing, O)] = U[-1, 1]$; *the voter's posterior on candidate j's policy preference upon observing a signal for moral alignment based on Bayes's rule.*
     - $\mu^*_v[t_j \neq t_v | (x_j, \varnothing)] = 1$; *the voter believes candidate j to be misaligned upon observing that the candidate campaigns on policy platforms.*
     - $\mu^*_v[t_j \neq t_v | (\varnothing, O)] = \frac{q}{1+q}$; *the voter's posterior on the probability of a candidate being morally misaligned given the campaign message $(\varnothing, O)$.*
2. ***Semi-Pooling***. *The following constitutes an equilibrium:*[29]

[29] Analytical solutions of the mixing probabilities α and β and the constant value $\tilde{B}$ do not appear obtainable in a reasonable time frame. Figure 3 in this section provides plots of numerical solutions of these variables for given values of λ and q, the exogenously determined parameters in the model.

- $\gamma^*_{A,I} = \begin{cases} \begin{cases} 1 & \textit{with prob. } \alpha \\ 0 & \textit{with prob. } 1-\alpha \end{cases} & \textit{if } |p_{A,I}| \in [0, \hat{\hat{p}}]; \hat{\hat{p}} \equiv \sqrt{\frac{1}{\lambda}(-\tilde{B} - \frac{1-\lambda}{1+\alpha})} \\ 0, & \textit{otherwise.} \end{cases}$

  where $\tilde{B}$ is the constant value voter receives from electing a candidate who sends either the aligned moral message or the policy platform that falls within the interval $[-\hat{\hat{p}}, \hat{\hat{p}}]$. A, I types mix between sending policy and moral messages if their policies fall within the interval.

- $\gamma^*_{D,I} = 1 \; \forall p_{D,I} \in P$; *$D, I$ types always campaign on policy platforms.*

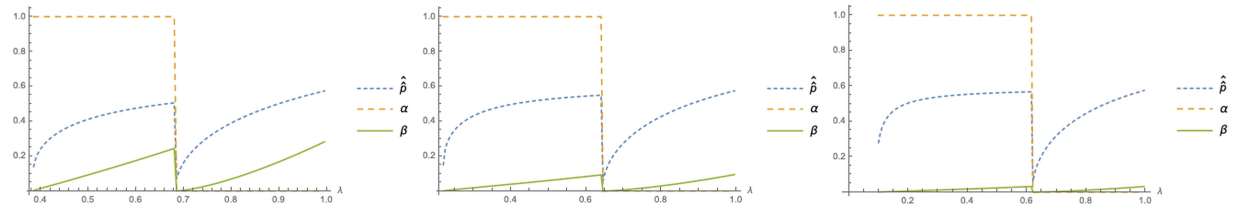
- $\gamma^*_{A,Z} = \gamma^*_{D,Z} = \begin{cases} 1 : (x^*_{A,Z}, m^*_{A,Z}) = (x^*_{D,Z}, m^*_{D,Z}) = (p'_Z, \varnothing) & \textit{with prob. } \beta \\ 0 & \textit{with prob. } 1-\beta \end{cases} \forall p_{A,Z}, p_{D,Z} \in P; |p'_Z| \in [0, \hat{\hat{p}}]$ *; Z types mix between sending policy and moral messages, and when sending policy messages, they select position p within the interval $[0, \hat{\hat{p}}]$ with probability $f(p)$.*

- $r^*_v = \begin{cases} 1 \quad \textit{if} \begin{cases} (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| < |x_2| \wedge |x_1| \in [0, \hat{\hat{p}}], |x_2| > \hat{\hat{p}} \\ (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| < |x_2| \wedge |x_1|, |x_2| > \hat{\hat{p}} \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (x_2, \varnothing), |x_2| > \hat{\hat{p}} \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (\varnothing, U) \end{cases} \\ \frac{1}{2}, \quad \textit{if} \begin{cases} (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1|, |x_2| \in [0, \hat{\hat{p}}] \\ (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (\varnothing, O) \wedge |x_1| \in [0, \hat{\hat{p}}] \\ (x_1, m_1) = (x_2, m_2) \end{cases} \\ 0, \quad \textit{otherwise.} \end{cases}$

- *Voter posterior beliefs:*
  - $\phi^*_v[(\varnothing, U)] = 1$; *the voter believes candidate j's policy preference to be the most extreme type upon observing that he reveals his moral misalignment.*
  - $\phi^*_v[(\varnothing, O)] = \begin{cases} U[-1, 1] & \textit{with prob. } \eta^*_v[Z|(\varnothing, O)] \\ U[-1, -\hat{\hat{p}}] + U[\hat{\hat{p}}, 1] & \textit{with prob. } 1 - \eta^*_v[Z|(\varnothing, O)] \end{cases}$; *the voter's posterior on candidate j's policy preference upon observing $(\varnothing, O)$.*
  - $\phi^*_v[(x_j, \varnothing), |x_j| \in [0, \hat{\hat{p}}]] = \begin{cases} U[-1, 1] & \textit{with prob. } \eta^*_v[Z|(x_j, \varnothing), |x_j| \in [0, \hat{\hat{p}}]] \\ x_j & \textit{with prob. } 1 - \eta^*_v[Z|(x_j, \varnothing), |x_j| \in [0, \hat{\hat{p}}]] \end{cases}$; *the voter's posterior on candidate j's policy preference upon observing a moderate policy announcement.*
  - $\eta^*_v[Z|(\varnothing, O)] = \frac{q(1-\beta)}{q(1-\beta) + \frac{1-q}{4}(\hat{p}(1-\alpha) + 1 - \hat{p})}$; *the voter's posterior on the likelihood of candidate j being a Z type upon observing $(\varnothing, O)$.*
  - $\eta^*_v[Z|(x_j, \varnothing), |x_j| \in [0, \hat{\hat{p}}]] = \frac{q\beta f(p)}{q\beta f(p) + \frac{1-q}{4}(1+\alpha)}$; *the voter's posterior on the likelihood of candidate j being a Z type upon observing moderate policy platform announcement $(x_j, \varnothing)$.*
  - $\mu^*_v[t_j \neq t_v|(\varnothing, O)] = \frac{q(1-\beta)}{2q(1-\beta) + \frac{(1-q)}{2}(\hat{p}(1-\alpha) + 1 - \hat{p})}$; *the voter's posterior on candidate's moral type being misaligned upon observing that the candidate cam-*

paigns on aligned moral messages.

- $\mu_v^*[t_j \neq t_v | x_j \in [0, \hat{\hat{p}}]] = \frac{q\beta f(p)+(1-q)}{2q\beta f(p)+\frac{(1-q)}{2}(1+\alpha)}$*; the voter's posterior on the likelihood of candidate* j *being morally misaligned upon observing that the candidate campaigns on policy platforms (in absolute terms) that fall in the interval* $[0, \hat{\hat{p}}]$
- $\mu_v^*[t_j \neq t_v | x_j \in (\hat{\hat{p}}, 1]] = 1$*; the voter believes candidate* j *to be morally misaligned upon observing that the candidate campaigns on policy platforms (in absolute terms) that fall in the interval* $(\hat{\hat{p}}, 1]$.



(a) $q = \frac{1}{4}$ (b) $q = \frac{1}{2}$ (c) $q = \frac{3}{4}$

Figure 3: Numerical Solutions of $\hat{\hat{p}}, \alpha, \beta$ for Various q

Figure 3 shows numerical solutions of the mixing probabilities and $\hat{\hat{p}}$ for three different values of q, and they show three interesting patterns. First, $\alpha \in \{0, 1\}$: A, I types do not mix. More interestingly, compared to their baseline counterparts, these candidates take actions that do not accommodate the voter's weight on alignment (e.g., they always choose to campaign on policies when their endowed policy is close enough to that of the voter when the latter cares more about moral alignment, and vice versa). The presence of Z types accounts for this behavior; A, I types consider it a best response to adopt such strategies for the sake of differentiation from Z types even if they become indistinguishable from D, I types by doing so.

This result shows that the insight from Callander and Wilkie (2007) who find that "the possibility that some candidates lie more than others affects the behavior of all candidates" remains to hold.[30] While D, I types' behavior is largely unaffected due to their severely limited options, the presence of Z types essentially flips A, I types' behavior, and the presence of infinite-cost types constrains cheap talkers' pandering behavior.

Second, as q goes up, the range of λ that can sustain the equilibrium increases (i.e., the lower bound on λ is decreasing in q). Lastly, Z types' mixing probability is generally

[30] More concretely, the introduction of zero-cost liars into the candidate pool induces "centrist" or moderate candidates to adopt truthful strategies, whereas they pool at the median voter's ideal point in the absence of cheap talkers.

increasing in λ just as in *Result 5*, but A, I types' behavior and q constrain the room for "pandering"; as q increases, β remains close to zero, which represents Z types pooling with A, I types with relatively extreme policy preferences. Why wouldn't they deviate to campaigning on policies?

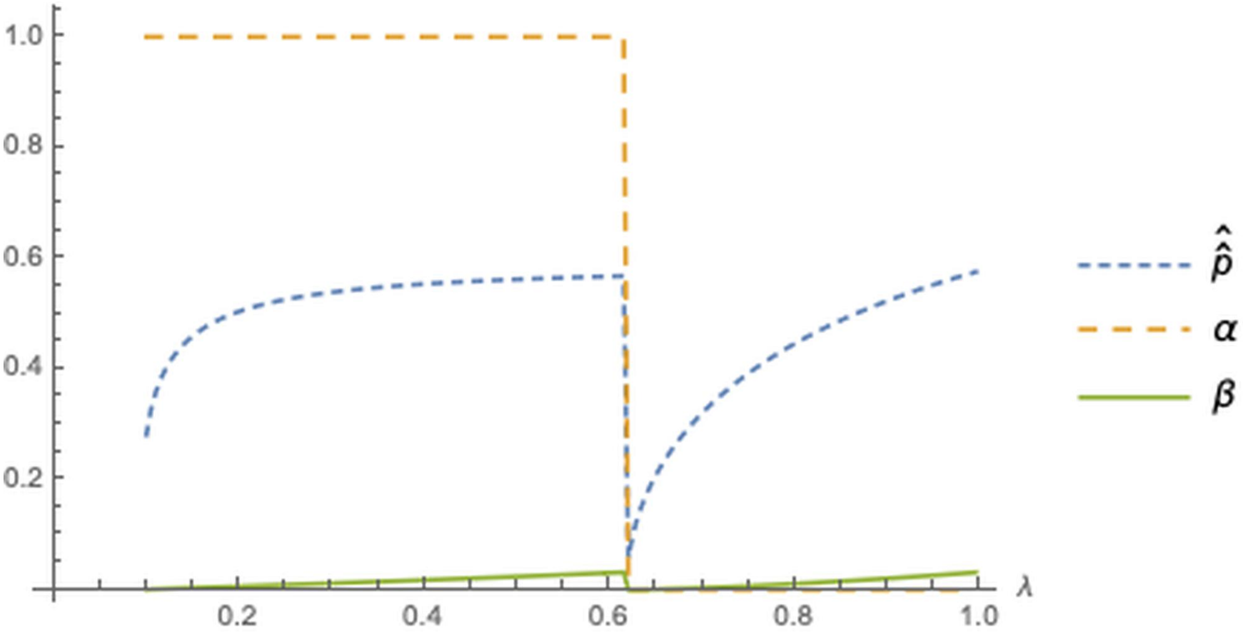


Figure 4: $q = \frac{3}{4}$

Consider figure 4. For the lower values of λ, campaigning on policies instead makes one indistinguishable not just from A, I types with moderate policy preferences, but also from D, I types. Since a low λ means that the voter cares much more about moral alignment, Z types find it a best response to appear as morally-aligned candidates (i.e., $\beta \sim 0$). For the higher values of λ, note that all A, I types are now campaigning on moral messages. As such, although the voter now cares more about policy alignment, she now considers one to be morally misaligned with a greater probability upon observing the policy platform in the campaign. Therefore, Z types would not find it profitable to deviate.

In sum, all results in this section suggest that being a cheap talker constitutes a significant advantage; in all equilibria, Z types are guaranteed at least one-half probability of winning the election. Second, the presence of liars significantly affects the behavior of A, I types. Conversely, the presence of infinite-cost types and the voter's (endogenously developed) preference for the honest types constrain Z types' pandering behavior.

5.4 Extension 3: Correlated Types and Cheap Talkers

This section analyzes an extended version where types are correlated and Z types are present. As the two corner cases and the separating equilibrium remain generally similar, it focuses on the semi-pooling equilibrium.

$\lambda \in (0,1)$**, Semi-Pooling Equilibrium**

The result below shows that some notable properties of each extension are preserved in this combined setting: first, the cut-points become asymmetric, just as in *Result 4*, with the left side of the cut-point (i.e., the policy spectrum less associated with communal values) being smaller in absolute terms compared to the right side cut-point. Second, A, I and Z types' mixing behaviors similar to those from *Result 6* continue to hold, although the magnitudes of mixing probabilities appear significantly different; A, I types actually mix, and Z types' mixing now follows the voter's weight parameter (λ) in a closer manner.

Result 7: *Suppose the voter cares about both policy and moral alignments ($\lambda \in (0,1)$). The semi-pooling equilibrium can be characterized as follows:*[31]

- $\gamma^*_{A,I} = \begin{cases} \begin{cases} 1 & \text{with prob. } \alpha \\ 0 & \text{with prob. } 1-\alpha \end{cases} & \text{if } p_{A,I} \in [\hat{p}_1, \hat{p}_2] \\ 0, & \text{otherwise.} \end{cases}$

  A, I types mix between sending policy and moral messages if their policies fall within the interval.

- $\gamma^*_{D,I} = 1 \ \forall p_{D,I} \in P$*; D, I types always campaign on policy platforms.*

- $\gamma^*_{A,Z} = \gamma^*_{D,Z} = \begin{cases} 1 : (x^*_{A,Z}, m^*_{A,Z}) = (x^*_{D,Z}, m^*_{D,Z}) = (p'_Z, \varnothing) & \text{with prob. } \beta \\ 0 & \text{with prob. } 1-\beta \end{cases} \forall p_{A,Z}, p_{D,Z} \in P; |p'_Z| \in [\hat{p}_1, \hat{p}_2]$ *; Z types mix between sending policy and moral messages, and when sending policy messages, they select position p within the interval $[\hat{p}_1, \hat{p}_2]$ with probability $f(p)$.*

[31] Analytical solutions of the mixing probabilities (α and β), the constant value $\hat{B}$, and the cut-points ($\hat{p}_1$ and $\hat{p}_2$) do not appear obtainable in a reasonable time frame. Figure 6 in this section provides plots of numerical solutions of these variables for given values of λ, q, and π, the exogenously determined parameters in the model.

- $r_v^* = \begin{cases} 1 \quad \textit{if} \begin{cases} (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| \in [\hat{p}_1, \hat{p}_2], x_2 \in [-1, \hat{p}_1) \vee (\hat{p}_2, 1] \\ (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge |x_1| < |x_2| \wedge x_1, x_2 \in [-1, \hat{p}_1) \vee (\hat{p}_2, 1] \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (x_2, \varnothing), x_2 \in [-1, \hat{p}_1) \vee (\hat{p}_2, 1] \\ (x_1, m_1) = (\varnothing, O), (x_2, m_2) = (\varnothing, U) \end{cases} \\ \frac{1}{2}, \quad \textit{if} \begin{cases} (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (x_2, \varnothing) \wedge x_1, x_2 \in [\hat{p}_1, \hat{p}_2] \\ (x_1, m_1) = (x_1, \varnothing), (x_2, m_2) = (\varnothing, O) \wedge x_1 \in [\hat{p}_1, \hat{p}_2] \\ (x_1, m_1) = (x_2, m_2) \end{cases} \\ 0, \quad \textit{otherwise.} \end{cases}$

- *Voter posterior beliefs:*

  - $\phi_v^*[(\varnothing, U)] = 1$*; the voter believes candidate j's policy type to be the worst possible type upon observing that he reveals his moral misalignment.*

  - $\phi_v^*[(\varnothing, O)] = \begin{cases} U[-1, 1] & \textit{with prob. } \eta_v^*[Z|(\varnothing, O)] \\ \begin{cases} U[\hat{p}_1, \hat{p}_2] & \textit{with prob. } \frac{\hat{p_2} - \hat{p_1}}{2} \\ U[-1, \hat{x_1}] & \textit{with prob. } \frac{1 + \hat{p_1}}{2} \\ U[\hat{p}_2, 1] & \textit{with prob. } \frac{1 - \hat{p_2}}{2} \end{cases} & \textit{with prob. } 1 - \eta_v^*[Z|(\varnothing, O)] \end{cases};$ *the voter's posterior on the candidate's policy preference upon observing* $(\varnothing, O)$*.*

  - $\phi_v^*[(x_j, \varnothing), x_j \in [\hat{p}_1, 0)] = \begin{cases} U[-1, 1] & \textit{with prob. } \eta_v^*[Z|(x_j, \varnothing), x_j \in [\hat{p}_1, 0)] \\ x_j & \textit{with prob. } 1 - \eta_v^*[Z|(x_j, \varnothing), x_j \in [\hat{p}_1, 0)] \end{cases};$ *the voter's posterior on the candidate's policy preference upon observing a moderate policy announcement in the interval* $[\hat{p}_1, 0)$*.*

  - $\phi_v^*[(x_j, \varnothing), x_j \in (0, \hat{p}_2]] = \begin{cases} U[-1, 1] & \textit{with prob. } \eta_v^*[Z|(x_j, \varnothing), x_j \in (0, \hat{p}_2]] \\ x_j & \textit{with prob. } 1 - \eta_v^*[Z|(x_j, \varnothing), x_j \in (0, \hat{p}_2]] \end{cases};$ *the voter's posterior on the candidate's policy preference upon observing a moderate policy announcement in the interval* $(0, \hat{p}_2]$*.*

  - $\phi_v^*[(x_j, \varnothing), x_j = 0] = \begin{cases} U[-1, 1] & \textit{with prob. } \eta_v^*[Z|(x_j, \varnothing), x_j = 0] \\ 0 & \textit{with prob. } 1 - \eta_v^*[Z|(x_j, \varnothing), x_j = 0] \end{cases};$ *the voter's posterior on the candidate's policy preference upon observing policy announcement 0.*

  - $\eta_v^*[Z|(\varnothing, O)] = \frac{q(1-\beta)}{q(1-\beta) + \frac{1-q}{2}(\frac{\hat{p_2} - \hat{p_1}}{2}(1-\alpha) + \frac{\hat{p_1} + 1 + 1 - \hat{p_2}}{2})}$*; the voter's posterior on the likelihood of candidate j being a Z type upon observing* $(\varnothing, O)$*.*

  - $\eta_v^*[Z|(x_j, \varnothing), x_j \in [\hat{p}_1, 0)] = \frac{q\beta f(p)}{q\beta f(p) + \frac{1-q}{2}(\pi + (1-\pi)\alpha)}$*; the voter's posterior on the likelihood of candidate j being a Z type upon observing a moderate policy platform in the interval* $[\hat{p}_1, 0)$*.*

  - $\eta_v^*[Z|(x_j, \varnothing), x_j \in (0, \hat{p}_2]] = \frac{q\beta f(p)}{q\beta f(p) + \frac{1-q}{2}(\pi\alpha + (1-\pi))}$*; the voter's posterior on the likelihood of candidate j being a Z type upon observing a moderate policy platform in the interval* $(0, \hat{p}_2]$*.*

- $\eta^*_v[Z|(x_j,\varnothing),x_j=0]=\frac{q\beta f(0)}{q\beta f(0)+\frac{1-q}{2}(\alpha+1)}$; *the voter's posterior on the likelihood of candidate j being a Z type upon observing policy platform 0.*
- $\mu^*_v[t_j\neq t_v|(\varnothing,O)]=\frac{q(1-\beta)}{2q(1-\beta)+(1-q)(\frac{\hat{p}_1+1+1-\hat{p}_2}{2})}$; *the voter's posterior on candidate's moral type being misaligned upon observing that the candidate campaigns on aligned moral messages.*
- $\mu^*_v[t_j\neq t_v|x_j\in[\hat{p}_1,0)]=\frac{q\beta f(p)+(1-q)\pi}{2q\beta f(p)+(1-q)((1-\pi)\alpha+\pi)}$; *the voter's posterior on candidate's moral type being misaligned upon observing that the candidate campaigns on a policy platform in the interval* $[\hat{p}_1,0)$.
- $\mu^*_v[t_j\neq t_v|x_j\in(0,\hat{p}_2]]=\frac{q\beta f(p)+(1-q)(1-\pi)}{2q\beta f(p)+(1-q)((1-\pi)+\pi\alpha)}$; *the voter's posterior on candidate's moral type being misaligned upon observing that the candidate campaigns on a policy platform in the interval* $(0,\hat{p}_2]$.
- $\mu^*_v[t_j\neq t_v|x_j=0]=\frac{q\beta f(0)+(1-q)}{2q\beta f(0)+(1-q)(1+\alpha)}$; *the voter's posterior on candidate's moral type being misaligned upon observing that the candidate campaigns on policy platform 0.*
- $\mu^*_v[t_j\neq t_v|x_j\in[-1,\hat{p}_1)\vee(\hat{p}_2,1]]=1$; *the voter believes that candidate j is morally misaligned with probability 1 with the candidate's policy announcement is in the interval* $[-1,\hat{p}_1)$ *or* $(\hat{p}_2,1]$.

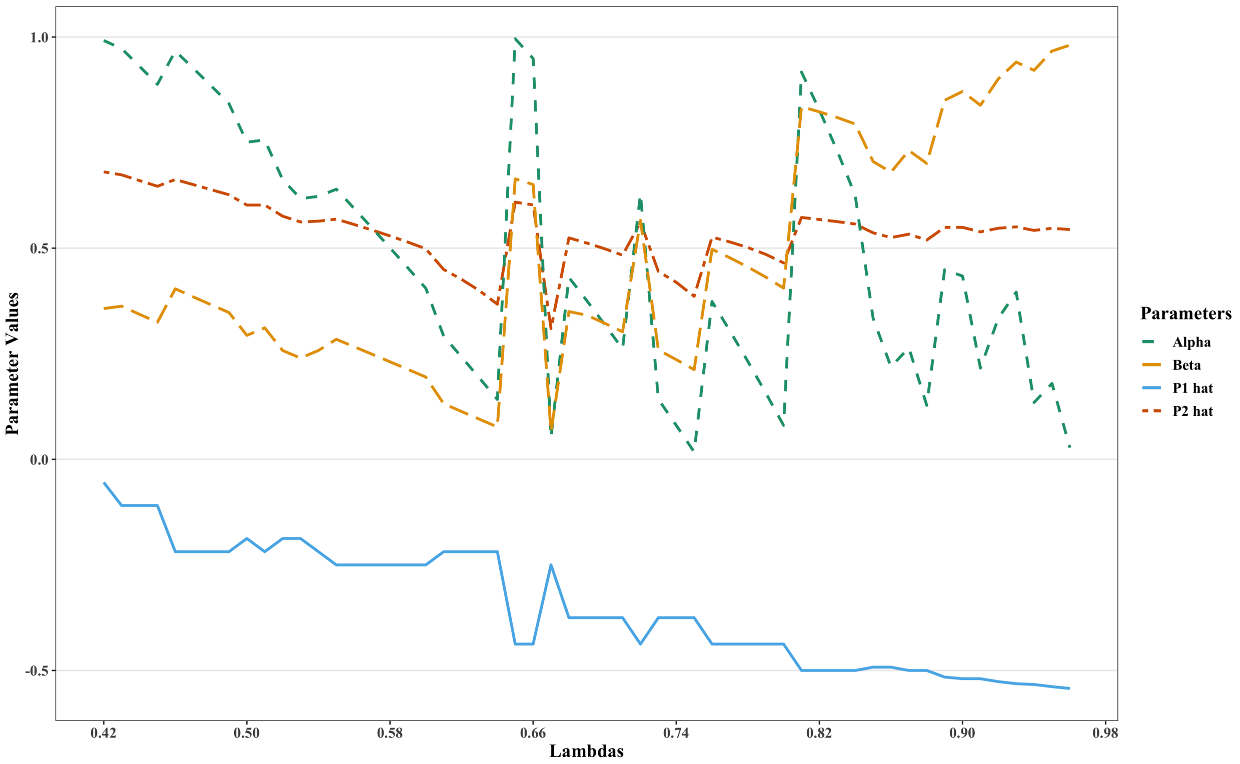


Figure 5: Numerical Solutions of $\hat{p}_1$, $\hat{p}_2$, α, β for $q=\frac{3}{4},\pi=\frac{2}{3}$

Figure 5 provides a set of numerical solutions of mixing probabilities and cut-points while assuming some fixed values of q and π,[32] and it shows that patterns from both extensions are generally preserved. First, $|\hat{p}_1| < |\hat{p}_2|$ remains to hold; A, I and Z types are less willing to campaign on policies less associated with the voter's values. Second, unlike *Result 6*, A, I types' mixing probability (α) is no longer binary. While it remains true that these candidates are still acting in the way that does not cater to the voter's weight parameter λ (e.g., they are more likely to campaign on policies when the voter cares more about moral alignment), they no longer campaign on moral messages with probability 1 for high values of λ. Finally, comparatively high values of β suggest that Z types now pander to the voter's taste in a closer manner (i.e., as $\lambda \to 1$, $\beta \to 1$). Again, their behavior appears to be constrained by A, I types' behavior, but they almost always campaign on policies as λ increases.

6 Discussion: Equilibrium Characteristics

Some notable patterns in candidate and voter behavior emerge across different versions of the model. This section discusses direct implications and potential empirical applications of these patterns where possible.

6.1 Candidate Behavior – Campaign Rhetoric

By construction, the formal analysis focuses on the type of campaign message selected by candidates. Their actions are largely determined not only by the voter and their own preferences, but also by the presence of cheap talkers and the correlation in types.

Divergence in Policy and Convergence in Moral Messages

Semi-pooling equilibria of *Results 4, 5*, and *7* all show that candidates diverge on moderate policies and converge on moral messages.[33] The divergence arises from the voter's preference for honest and moderate candidates, as it prevents cheap talkers from pooling at the voter's idea point if they campaign on policy platforms. On the moral front, candidates who campaign on moral messages converge because revealing oneself to be morally misaligned is often strictly dominated; even morally-misaligned candidates can improve their prospect by campaigning on policy platforms instead in most settings.[34]

[32] A three set of figures that assume different values of q are provided in the Appendix. Due to the long computation time required for obtaining one set of solutions for a given value of λ, the current section omits a little more detailed comparative statics.

[33] A small exception: morally-misaligned D, I types do campaign on extreme policies too. Regardless, they are not expected to campaign on moral messages that signal themselves to be of an opposite type, so the point on the convergence in moral messages holds.

[34] Note that the exact convergence results from the modeling choice that restricts the moral type space to

In relation to Callander and Wilkie (2007) and Kartik and McAfee (2007) who also predict divergence in policies, the results from the current model can be construed as an extension that not only predicts policy divergence widely observed in real world elections, but also predicts the use of language that signals moral alignment with the voter by candidates if they do campaign on moral messages. This constitutes a prediction that is empirically testable.

Correlation Constrains Policy Platform Announcements

Allowing an individual's moral identity to be correlated with one's policy preferences yields an intuitive yet important prediction that candidates campaign less on moderate policies associated with the opposite moral values. More specifically, semi-pooling equilibria of *Results 4* and *7* show that the interval over which morally-aligned candidates are willing to campaign on policy platforms is asymmetric, with the side better associated with the voter's moral type being larger (e.g., the conservative side of the spectrum if the voter's moral type is communal).

What would this mean in the real world? Consider, for instance, a Republican candidate with communal values and policy preference of imposing a moderately higher tax rate for the rich. Albeit moderate and actually close to the voter's ideal point, his/her campaign speech might be filled with moral messages that signal moral alignment to minimize the possibility of being perceived as a morally-misaligned candidate. At a high level, an empirical extension could test this claim by analyzing campaign speeches delivered by candidates whose policy preference is slightly moderate but more liberal (conservative) than their conservative (liberal) electorate.

Election Outcomes: Liars are Advantaged but Constrained

Cheap talkers win with probability at least $\frac{1}{2}$ in all equilibria whenever they are present, so they are advantaged. That said, honest and morally-advantaged (A, I type) candidates can do just as well when their policy preferences are sufficiently moderate. Moreover, as shown in semi-pooling equilibria of *Results 6* and *7*, the presence of infinite-cost types and the voter's preference for such honest candidates constrain Z types' pandering behavior. For example, Z types' mixing probability for campaigning on moderate policies, β, does not monotonically

be discrete. That said, candidates presumably adopt languages that signal similar values to the voter, consequently making them difficult to set apart in terms of moral values. Put another way, their degree of being "communal" or "universal" may not be as finely distinguishable as their policy positions on tax rates or healthcare.

increase as λ approaches 1.

The concept of zero-cost types seems nearly impossible to apply to the real-world in a literal sense because the psychological costs candidates incur from lying are not readily observable. Nevertheless, if some factual and policy inconsistencies found among politicians can be construed as a proxy for their tendency to lie, then this model provides an account for why and how such liars can be selected by the voter in the real world.

6.2 Voter Behavior

Like other models of electoral competition, driving the strategic behavior of candidates is voter preference and learning. With the addition of heterogeneous types over the moral dimension and the presence of cheap talkers, the cognitive demands on the voter in the model are significant, hence might appear rather unrealistic. Nevertheless, the voting behavior (i.e., candidate selection) predicted by the model seems generally consistent and plausible. Below discusses two such behaviors that pertain to extremists.

Asymmetric Learning Supplemented by Correlation

A candidate's type affects the amount of information delivered to the voter during the campaign. For example, in semi-pooling equilibria of *Results 4, 6,* and *7*, extreme policy announcements fully reveal a candidate to be of a honest but morally-misaligned type. In addition, the correlation in types allows the voter to assign a greater (or lower) probability on candidate's moral and policy preferences. Consequently, the voter prefers a candidate with a relatively extreme policy preference but on the "right" side over a candidate with a moderate policy but on the "wrong" side. This in turn results in the candidate behavior of campaigning less on "opposite" policy platforms.

The caveat of the policy announcements having to be generally moderate aside, the prediction on the voter's preference calls for an empirical application. If found to hold, this theoretical result could account for the phenomenon that is inconsistent with the one-dimensional median voter theorem that predicts voters to favor candidates who are closer to their ideal points.[35]

Moral Messages as a Preferable Substitute for Policy Platforms

[35] Hirsch and Shotts (2015) also find a similar result in their model of policy development.

Results 6 and *7* show that the presence of cheap talkers induces candidates to behave in ways that consequently leave the voter indifferent between moderate policy platforms and moral announcements. As a result, the voter selects morally-aligned candidates with at least $\frac{1}{2}$ probability whenever candidates signal moral alignment.

A substantive implication from this appeal of moral messages as a substitute for the policy platform is that policy-wise extreme but morally-aligned candidates can win with a high probability just by signalling their moral alignment. While existing empirical works on extremists (see e.g., Hall 2015; Hall and Thompson 2018) show that they are disadvantaged in general elections, some ideologically extreme candidates do win offices. Broockman et al. (2018)'s recent finding that local party leaders prefer "nominating candidates who are similar to typical co-partisans, not centrists" could be related to this theoretical result; these leaders might prefer such extreme candidates because they are morally aligned with co-partisans. This theoretical result, then, provides another possible explanation for how and why extreme candidates can win the election.

7 Conclusion

Policy platforms and party identity are not the only aspect of candidates voters consider in elections. Based on empirical regularities found in moral psychology and political science, this paper develops a theory of electoral competition with heterogeneity in individuals' party and moral identity. The special features of the model include the correlation in types – specifically, one's moral identity affects party affiliation – and the presence of liars who can announce to be of moral or party type different from their true type. Analyzing the model reveals that candidates generally diverge on policy but converge on moral messages that signal their alignment with the voter. In addition, the correlation in types makes candidates unwilling to campaign on policies less associated with the voter's values. This implies that morally aligned but extremely partisan candidates have a significant chance of winning. Similarly, when cheap talkers are present in the model, candidates behave in a way that leaves the voter indifferent between moderate partisan and moral alignment. This in turn allows such morally-aligned extreme candidates to win with a high probability. In sum, this theory provides two alternative mechanisms through which extremists can win political offices, both of which do not appear well-explored in the literature.

While some aspects of the model including candidates' tendency to lie might be difficult to observe, some predictions on candidate and voter behavior appear empirically testable by analyzing candidate speech and voting data similar to those showcased in Enke (2018). On the theoretical front, the current framework lacks the intermediate actors (e.g., media, political activists, etc.) who can influence the weight of importance voters assign to policy or moral values (i.e., the parameter λ in the current model). Extending the model based on previous work such as Besley and Prat (2006) could be profitable. More broadly, this interdisciplinary approach could be applied to other settings where political actors have an incentive to influence audience with heterogeneous beliefs or preferences.

References

Adams, James et al. (2011). “When Candidates Value Good Character: A Spatial Model with Applications to Congressional Elections”. en. In: *The Journal of Politics* 73.1, pp. 17–30.

Annenberg Public Policy Center, The (2014). *Americans know surprisingly little about their government, survey finds.* en-US.

Ansolabehere, Stephen and James M. Snyder (2002). “The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices, 1942–2000”. In: *Election Law Journal: Rules, Politics, and Policy* 1.3, pp. 315–338.

Bailenson, J. N. et al. (2008). “Facial Similarity between Voters and Candidates Causes Influence”. en. In: *Public Opinion Quarterly* 72.5, pp. 935–961.

Baldassarri, Delia and Andrew Gelman (2008). “Partisans without Constraint: Political Polarization and Trends in American Public Opinion”. eng. In: *AJS; American journal of sociology* 114.2, pp. 408–446.

Banks, Jeffrey S (1990). “A Model of Electoral Competition with Incomplete Information”. In: *Journal of Economic Theory* 50.2, pp. 309–325.

Bernhardt, Dan, Odilon Câmara, and Francesco Squintani (2011). “Competence and Ideology”. en. In: *The Review of Economic Studies* 78.2, pp. 487–522.

Besley, Timothy and Torsten Persson (2019). “Democratic Values and Institutions”. en. In: *American Economic Review: Insights.*

Besley, Timothy and Andrea Prat (2006). “Handcuffs for the Grabbing Hand? Media Capture and Government Accountability”. en. In: *American Economic Review* 96.3, pp. 720–736.

Bonica, Adam and Gary W. Cox (2018). “Ideological Extremists in the U.S. Congress: Out of Step but Still in Office”. English. In: *Quarterly Journal of Political Science* 13.2, pp. 207–236.

Broockman, David E. (2016). “Approaches to Studying Policy Representation: Studying Policy Representation”. en. In: *Legislative Studies Quarterly* 41.1, pp. 181–215.

Broockman, David E et al. (2018). “Having Their Cake and Eating It, Too: Why Local Party Leaders Don’t Support Nominating Centrists”. en. In: *British Journal of Political Science* Forthcoming, pp. 1–106.

Callander, Steven (2008). “Political Motivations”. In: *The Review of Economic Studies* 75.3, pp. 671–697.

Callander, Steven and Simon Wilkie (2007). “Lies, damned lies, and political campaigns”. In: *Games and Economic Behavior* 60.2, pp. 262–286.

Calvert, Randall L. (1985). “Robustness of the Multidimensional Voting Model: Candidate Motivations, Uncertainty, and Convergence”. In: *American Journal of Political Science* 29.1, pp. 69–95.

Caprara, Gian Vittorio et al. (2006). “Personality and Politics: Values, Traits, and Political Choice”. en. In: *Political Psychology* 27.1, pp. 1–28.

Converse, Philipe E. (1964). “The Nature of Belief Systems in Mass Publics”. In: *Ideology and Discontent.* David E. Apter. New York: The Free Press, pp. 206–261.

Deason, Grace and Marti Hope Gonzales (2012). “Moral Politics in the 2008 Presidential Convention Acceptance Speeches”. In: *Basic and Applied Social Psychology* 34.3, pp. 254–268.

Downs, Anthony (1957). *An Economic Theory of Democracy.* English. 1st edition. Boston: Harper and Row.

Enke, Benjamin (2017). *Kinship, Cooperation, and the Evolution of Moral Systems.* Working Paper 23499. National Bureau of Economic Research.

— (2018). *Moral Values and Voting.* Working Paper 24268. National Bureau of Economic Research.

Festinger, Leon (1962). *A Theory of Cognitive Dissonance.* en. Stanford University Press.

Fu, Feng et al. (2012). "The Evolution of Homophily". en. In: *Scientific Reports* 2, p. 845.

Graham, Jesse, Jonathan Haidt, and Brian A. Nosek (2009). "Liberals and conservatives rely on different sets of moral foundations." en. In: *Journal of Personality and Social Psychology* 96.5, pp. 1029–1046.

Groseclose, Tim (2001). "A Model of Candidate Location When One Candidate Has a Valence Advantage". In: *American Journal of Political Science* 45.4, pp. 862–886.

Haidt, Jonathan (2013). *The Righteous Mind: Why Good People Are Divided by Politics and Religion.* English. Reprint edition. New York, N.Y: Vintage.

Hall, Andrew B. (2015). "What Happens When Extremists Win Primaries?" en. In: *American Political Science Review* 109.1, pp. 18–42.

Hall, Andrew B. and Daniel M. Thompson (2018). "Who Punishes Extremist Nominees? Candidate Ideology and Turning Out the Base in US Elections". en. In: *American Political Science Review* 112.3, pp. 509–524.

Henrich, Joseph (2015). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter.* English. Princeton: Princeton University Press.

Hirsch, Alexander V. and Kenneth W. Shotts (2015). "Competitive Policy Development". en. In: *American Economic Review* 105.4, pp. 1646–1664.

Hotelling, Harold (1929). "Stability in Competition". In: *The Economic Journal* 39.153, pp. 41–57.

Huber, Gregory A. and Neil Malhotra (2016). "Political Homophily in Social Relationships: Evidence from Online Dating Behavior". In: *The Journal of Politics* 79.1, pp. 269–283.

Iyengar, Shanto (2005). "Speaking of Values: The Framing of American Politics". In: *The Forum* 3.3.

Kartik, Navin and R. Preston McAfee (2007). "Signaling Character in Electoral Competition". en. In: *American Economic Review* 97.3, pp. 852–870.

Kinder, Donald R. and Nathan P. Kalmoe (2017). *Neither Liberal nor Conservative: Ideological Innocence in the American Public.* English. 1 edition. Chicago ; London: University of Chicago Press.

Kinder, Donald R. and David O. Sears (1985). "Public Opinion and Political Action". In: *The Handbook of Social Psychology.* G. Lindzey and E. Aronson. Vol. 2. New York: Random House, pp. 659–741.

Lakoff, George (2002). *Moral Politics : How Liberals and Conservatives Think.* English. Second edition. Chicago: University Of Chicago Press.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a Feather: Homophily in Social Networks". In: *Annual Review of Sociology* 27.1, pp. 415–444.

Meirowitz, Adam (2007). "Probabilistic Voting and Accountability in Elections with Uncertain Policy Constraints". en. In: *Journal of Public Economic Theory* 9.1, pp. 41–68.

Ortoleva, Pietro and Erik Snowberg (2015). “Overconfidence in Political Behavior”. en. In: *American Economic Review* 105.2, pp. 504–535.

Piurko, Yuval, Shalom H. Schwartz, and Eldad Davidov (2011). “Basic Personal Values and the Meaning of Left-Right Political Orientations in 20 Countries”. en. In: *Political Psychology* 32.4, pp. 537–561.

Ryan, Timothy J. (2014). “Reconsidering Moral Issues in Politics”. In: *The Journal of Politics* 76.2, pp. 380–397.

Schwartz, Shalom H., Gian Vittorio Caprara, and Michele Vecchione (2010). “Basic Personal Values, Core Political Values, and Voting: A Longitudinal Analysis”. en. In: *Political Psychology* 31.3, pp. 421–452.

Sherman, Ryne A. (2018). “Personal values and support for Donald Trump during the 2016 US presidential primary”. In: *Personality and Individual Differences* 128, pp. 33–38.

Somin, Ilya (2014). *What No One Talks About During Election Season: Voter Ignorance*. en.

Stone, Walter J. and Elizabeth N. Simas (2010). “Candidate Valence and Ideological Positions in U.S. House Elections”. en. In: *American Journal of Political Science* 54.2, pp. 371–388.

Appendix

This section provides the formal proofs of results presented in the main text of the paper. As a reminder, definitions and notations are as follows:

1. Individual i moral type: $t_i \in \{O, U\}$, $Pr(t_i = O) = Pr(t_i = U) = \frac{1}{2}$.

2. Individual i policy preferences: $p_i \in U[-1, 1]$, drawn iid for the baseline model.

3. Candidate j's action: (x_j, m_j) represents policy and moral type announcements, respectively.

4. Candidate j's strategy: $\gamma_j \in \{0, 1\}$; $\gamma_j = 1$ if the candidate announces his policy platforms, 0, otherwise.

5. Voter i's strategy: $r_i((x_1, m_1), (x_2, m_2))$ as the probability that voter i selects candidate 1 over 2.

6. Voter posterior beliefs on candidate j's policy (moral) preference upon seeing moral (policy) preference:

   - Policy preference (distribution): $\phi[(x_j, m_j) = (\varnothing, t_j)]$.
   - Moral misalignment: $\mu[t_j \neq t_i | (x_j, m_j)]$.
   - Probability of being a Z-type: $\eta[Z_j = 1 | (x_j, m_j)]$.

7. The baseline model includes two types of candidates – A for "morally-advantaged" and D for "morally-disadvantaged" – which depends on the moral type of the voter. In this construction, then, candidate with moral type $\{O\}$ has an advantage, as his moral type corresponds to that of the representative voter.

8. The extension that correlates the types makes an individual i's policy preference to be conditional on her moral type: $Pr(p_i \in [-1, 0] | t_i = U) = Pr(p_i \in [0, 1] | t_i = O) = \pi > \frac{1}{2}$.

9. The extension that introduces the zero-cost types adopt the following notations: Z-types for the zero-cost types and I for infinite-cost (honest) types who continue to suffer infinite disutility from lying.

The main solution concept is the perfect Bayesian equilibrium for all proofs.

A Baseline Model Proofs

A.1 Case 1: $\lambda = 0$

Note that the voter does not care about policy alignment at all. Then, conjecture an equilibrium where upon seeing a policy platform announced, the voter believes the given candidate's moral value to not match hers:

$$\mu[t_j \neq t_v | (x_j, m_j) = (p_j, \varnothing)] = 1$$

Given such a belief, the morally-advantaged candidate (A) always announces his moral type; doing otherwise gets him labeled as the misaligned type, hence constitutes a dominated strategy. For the disadvantaged type (A), either action fully reveals his moral misalignment, hence indifferent. The characterization of the equilibrium is reported in the text.

A.2 Case 2: $\lambda = 1$

Note that the voter does not care about moral alignment at all. Just as in the case above, the voter can infer something about the candidate's policy preference upon seeing that the candidate chose to reveal his moral type even when he knows that the voter does not care.

Lemma 1: There does not exist an equilibrium where candidates adopt a cut-point strategy in which a candidate with policy preference $|p_j| \in [0, \bar{p}]$ announces his policy preference, and moral type, otherwise.

Proof: Suppose not, and conjecture such an equilibrium. Voter's posterior belief on a candidate who does not reveal his policy preference becomes:

$$\phi[p_j|(x_j, m_j) = (\varnothing, t_j)] = \frac{\bar{p}+1}{2}$$

But given such a belief, any candidate j with policy preference $p'_j \in (\bar{p}, \frac{\bar{p}+1}{2})$ can profitably deviate by announcing his policy preference p'_j instead of moral type:

$$E[u_v|(x_j, m_j) = (\varnothing, t_j)] = -(\frac{\bar{p}+1}{2} + \frac{(1-\bar{p})^2}{12})$$

$$E[u_v|(x_j, m_j) = (p'_j, \varnothing)] = -(p'_j)^2$$

By construction, $\left|p'_j\right| < \left|\frac{1+\bar{p}}{2}\right|$, which means $E[u_v|(\varnothing, t_j)] < E[u_v|(p'_j, \varnothing)]$. Therefore, the candidate prefers to announce his policy preference instead, thereby violating the conjectured equilibrium behavior. As this applies to both types of moral values, no candidate under this setting can adopt such a strategy, and we are done. □

With a cut-point strategy ruled out, the best either candidate can do is to campaign on policy platforms. To prevent profitable deviation by candidates, the voter can form the following posterior upon observing moral messages from candidate j:

$$\phi[|p_j||(x_j, m_j) = (\varnothing, t_j)] = 1$$

That is, assume candidate j to be the worst possible policy type when he exhibits an off-path behavior (i.e., sending moral messages). The characterization of the equilibrium is reported in the text.

A.3 Case 3: $\lambda \in (0,1)$

Note that the voter cares about both policy and moral alignments.

Equilibrium 1 – Separating:
First conjecture an equilibrium where for some values of λ morally-advantaged candidate only reveals his moral type $\forall p_A \in [-1,1]$ and the disadvantaged types only reveal their policy preferences. Then, the voter's posterior becomes:

- On disadvantaged type's policy (off-path): $\phi[(\varnothing, \{U\})] = 1$
- On advantaged type's policy: $\phi[(\varnothing, \{O\})] = U[-1,1]$
- $\mu[t_j \neq t_v | m_j \neq t_v] = 1$

That is, the voter assumes the policy preference of a morally disadvantaged candidate to hold the worst possible policy preference if the latter reveals his moral type. For the advantaged type who is always expected to reveal his moral type, the voter takes the expected value for his policy preference based on a predefined uniform distribution $f(\cdot)$. Finally, if a candidate does not reveal his moral type or reveals to be an opposite type, the voter believes that the candidate holds the opposite moral value with probability 1.

Now consider each candidate type's incentive to deviate. First, consider D candidates, the disadvantaged type. Each of their action results in following expected utilities for the voter:

$$E[u_v|(\varnothing, \{U\})] = -\lambda - (1-\lambda)$$

$$E[u_v|(p_D, \varnothing)] = -\lambda p_D^2 - (1-\lambda)$$

Since $|p_D| \in [0,1]$, $E[u_v|(p_D, \varnothing)] \geq E[u_v|(\varnothing, \{U\})]$, hence no incentive to deviate by revealing his moral type. Now consider candidate type A's incentive to deviate:

$$E[u_v|(\varnothing, \{O\})] = -\lambda[\phi(p_A^2)] - 0 = -\lambda\sigma^2 = -\lambda\frac{1}{3}$$

$$E[u_v|(p_A, \varnothing)] = -\lambda p_A^2 - (1-\lambda)$$

For candidate type A to never find it profitable to deviate, consider the case when the candidates have the best possible policy preference (i.e., $p_A = p_v = 0$):

$$-\lambda\sigma^2 > -\lambda * 0 - (1-\lambda) \Leftrightarrow \lambda < \frac{1}{1+\sigma^2} = \frac{3}{4} \equiv \underline{\lambda}$$

The characterization of the equilibrium assuming $\lambda < \underline{\lambda}$ is reported in the text.

Equilibrium 2 – Semi-Pooling:
Conjecture an equilibrium where some of the advantaged candidates prefers to announce their policy preference (i.e., they adopt a cut-point strategy where candidates with policy

preferences below some cut-point $|\bar{p}|$ announce their policy types, and moral types, otherwise). Suppose that the disadvantaged candidates only reveal their policy types $\forall p_D \in P$. Then, the voter's posteriors become:

- $\phi[(\varnothing, \{U\})] = 1$
- $\phi[(\varnothing, \{O\})] = U[-1, -\bar{p}] + U[\bar{p}, 1]$
- $\mu[t_j \neq t_v | (p_j, \varnothing); p_j \in [0, \bar{p}]] = \frac{1}{2}$
- $\mu[t_j \neq t_v | (p_j, \varnothing); p_j \in (\bar{p}, 1]] = 1$

The first belief is same as that of the equilibrium 1 above. Voter's belief on morally-aligned candidate's policy preference given the announcement of his moral type is the expected value based on the truncated distribution $U[-1, \bar{p}) + U(\bar{p}, 1]$. Accordingly, the third belief states that upon observing the policy that falls in the range $[-\bar{p}, \bar{p}]$, voter considers the candidate to be equally likely to be morally-misaligned, since both types of candidates are conjectured to announce policies. Finally, voter knows with probability 1 that a candidate is morally-misaligned if the latter announces the policy whose absolute value falls within the $(\bar{p}, 1]$ interval.

Now consider each type of candidate's incentive to deviate. Candidate type D does not have any incentive to deviate by the same reasoning as above in equilibrium 1. (Indeed, he now has a better reason to campaign only on policies if his policies are below the cut-point.) Consider candidate type A's incentive to deviate:

$$E[u_v | (\varnothing, \{O\})] = -\lambda[\phi(p_A^2)] - 0 = -\lambda((\frac{1+\bar{p}}{2})^2 + \frac{1}{12}(1-\bar{p})^2)$$

$$E[u_v | (p_A, \varnothing)] = -\lambda p_A^2 - \frac{(1-\lambda)}{2}$$

The cut-point $\bar{p}$ satisfies:

$$-\lambda((\frac{1+\bar{p}}{2})^2 + \frac{1}{12}(1-\bar{p})^2) = -\lambda(\bar{p})^2 - \frac{(1-\lambda)}{2} \Leftrightarrow$$

$$|\bar{p}| \equiv \frac{1}{4}(1 \pm \frac{\sqrt{3}\sqrt{7\lambda - 4}}{\sqrt{\lambda}})$$

And λ must satisfy the following to ensure that $|\bar{p}| \in [0, 1]$:[36]

$$\lambda > \frac{4}{7} \equiv \bar{\lambda}$$

Now consider the advantaged type's incentive to deviate when his policy type is relatively close to the cut-point (i.e., $p_A = \bar{p} + \epsilon$):

$$E[u_v | (\varnothing, \{O\})] = -\lambda((\frac{1+\bar{p}}{2})^2 + \frac{1}{12}(1-\bar{p})^2)$$

[36] $\bar{\lambda} \equiv \frac{3}{3D^2+4}$ for a more general uniform distribution $[-D, D]$.

$$E[u_v|(p_A, \varnothing)] = -\lambda(\bar{p} + \epsilon)^2 - (1 - \lambda)$$

For this candidate to not deviate, it must be that $E[u_v|(\varnothing, \{O\})] \geq E[u_v|(p_A, \varnothing)]$. But since $-\lambda(\bar{p} + \epsilon)^2 < -\lambda(\bar{p})^2$ and $-(1 - \lambda) < \frac{(1-\lambda)}{2}$, the inequality must hold. Therefore, such a candidate will not find it profitable to deviate. A symmetric line of reasoning applies for the candidate with policy preference $p_A = \bar{p} - \epsilon$. The characterization of the equilibrium is reported in the text.

Claim: There does not exist an equilibrium where both moral types adopt cut-point strategies.

Proof: Suppose not; denote candidate type D's cut-point as p_D^* and p_A^* for candidate type A. Since the latter has a moral advantage, the point at which he becomes indifferent between announcing his moral type and policy preference will be lower in absolute terms (i.e., closer to 0, the voter's ideal point) than that of candidate A:

$$|p_D^*| > |p_A^*|$$

Then, in such an equilibrium, the voter forms following posterior beliefs on candidate D:

- $\phi((\varnothing, \{U\})) = U[-1, -p_D^*] + U[p_D^*, 1]$
- $\mu[t_A \neq t_v|(p_D, \varnothing); p_D \in [p_A^*, p_D^*]] = 1$

Given such beliefs, consider candidate A's incentive to deviate when his preference is $p_D = p_D^* + \epsilon$:

$$E[u_v|(\varnothing, \{U\})] = -\lambda[\phi(p_D^2)] - 0 = -\lambda((\frac{1 + p_D^*}{2})^2 + \frac{1}{12}(1 - p_D^*)^2) - (1 - \lambda)$$

$$E[u_v|(p_D, \varnothing)] = -\lambda(p_D^* + \epsilon)^2 - (1 - \lambda)$$

With sufficiently small ϵ, such candidate A is strictly better off making his policy preference known (i.e., $E[u_v|(p_D, \varnothing)] > E[u_v|(\varnothing, \{U\})]$), which constitutes a profitable deviation. Therefore, such an equilibrium cannot be sustained, and we are done. □

From derivations of two equilibria above, it becomes clear that $\underline{\lambda} = \frac{3}{4} > \bar{\lambda} = \frac{4}{7}$. Therefore, $\forall \lambda \in (\frac{4}{7}, \frac{3}{4})$, both equilibria can be sustained.

Existence of Other Equilibria

Now we check the existence of other equilibria. This proof shows that there are at least two other equilibria that share the same characterizations as *result 1* and *result 2*.

First, conjecture an equilibrium similar to that of *result 1* while still assuming $\lambda \in (0, 1)$; all types are conjectured to campaign on moral messages regardless of their policy preferences. Then, similar to case 1, the voter would form the following posterior upon observing policy platform announcements from candidate:

$$\mu[t_j \neq t_v|(p_j, \varnothing)] = 1$$

Then, following the logic from case 1, the A types would consider it a best response to always campaign on moral messages, while the D types are indifferent between either action, as their misalignment is bound to be discovered. As a result, for all values of λ, the same exact characterization as *result 1* applies. □

Now conjecture an equilibrium similar to that of *result 2* while still assuming $\lambda \in (0,1)$; all types are conjectured to campaign on policy platforms regardless of their moral alignments. Based on this initial conjecture, define voters' posteriors on the off-path behavior (i.e., campaigning on moral messages $(\varnothing, t_j)$) as follows:

$$\mu[t_j \neq t_v | (\varnothing, t_j)] = 1$$

$$\phi[(\varnothing, t_j)] = 1$$

That is, whenever the voter observes a moral message, she assumes the candidate to be misaligned with probability 1 and to hold the most extreme policy preference. Then,

$$E[u_v | (\varnothing, t_j)] = -\lambda - (1-\lambda)$$

As this constitutes the lowest possible expected utility, deviating to campaigning on moral messages is dominated for both types of candidates. With the only change from *result 2* being $\lambda \in (0,1)$, the same equilibrium characterization applies. □

B Proof of Proposition 1: Welfare Comparison

The result above shows that there can be multiple equilibria when $\lambda \in (\frac{4}{7}, \frac{3}{4})$, which raises a question on the voter's expected welfare under each equilibrium. First, given the equal prior probability of moral type realization, there is $\frac{1}{4}$ chance that both candidates are of the disadvantaged type and probability $\frac{3}{4}$ that at least one advantaged type is in the election.

Suppose $\lambda \in (\frac{4}{7}, \frac{3}{4})$. Then, under the separating equilibrium, the voter's expected utility is as follows:

$$E[u_v|SE] = \frac{1}{4}[-\lambda E[\phi(p_D^2)|(p_D, \varnothing)]] - (1-\lambda)] + \frac{3}{4}[-\lambda\sigma^2]$$

Now consider her expected utility under the semi-pooling equilibrium:

$$E[u_v|SPE] = \frac{1}{4}[-\lambda E[\phi(p_D^2)|(p_D, \varnothing)]] - (1-\lambda)] +$$

$$\frac{1}{2}[(\int_0^{\bar{p}} f(p_A)dp_A)(\int_0^{\bar{p}} f(p_D)dp_D)[-\lambda E[\phi(p_j^2)|p_j \in [0,\bar{p}]] - \frac{(1-\lambda)}{2}] +$$

$$(\int_{\bar{p}}^1 f(p_A)dp_A)(\int_0^{\bar{p}} f(p_D)dp_D)[-\lambda E[\phi(p_D^2)|p_D \in [0,\bar{p}]] - (1-\lambda)] +$$

$$(\int_0^{\bar{p}} f(p_A)dp_A)(\int_{\bar{p}}^1 f(p_D)dp_D)[-\lambda E[\phi(p_A^2)|p_A \in [0,\bar{p}]] - 0] +$$

$$(\int_{\bar{p}}^{1} f(p_A)dp_A)(\int_{\bar{p}}^{1} f(p_D)dp_D)[-\lambda E[\phi(p_A^2)|(\varnothing,\{O\})]]]+$$

$$\frac{1}{4}[(\int_{0}^{\bar{p}} f(p_A)dp_A)^2[-\lambda E[\phi(p_A^2)|p_A \in [0,\bar{p}]] - 0]+$$

$$2(\int_{0}^{\bar{p}} f(p_A)dp_A)(\int_{\bar{p}}^{1} f(p_A)dp_A)[-\lambda E[\phi(p_A^2)|p_A \in [0,\bar{p}]] - 0]+$$

$$(\int_{\bar{p}}^{1} f(p_A)dp_A)^2[-\lambda E[\phi(p_A^2)|(\varnothing,\{O\})]]]]$$

Still assuming the preference distribution to be $U[-1,1]$, the expression simplifies to:

$$E[u_v|SPE] = \frac{1}{4}[-\lambda E[\phi(p_D^2)|(p_D,\varnothing)]] - (1-\lambda)] - \frac{\lambda}{4}\{(\bar{p})^3 - \bar{p} + 1\} - (1-\lambda)\frac{\bar{p}}{2}(1 - \frac{(\bar{p})^2}{2})$$

Then, comparing expected utilities from the two equilibria boils down to the following inequality:

$$-\frac{\lambda}{4} > -\frac{\lambda}{4}\{(\bar{p})^3 - \bar{p} + 1\} - (1-\lambda)\frac{\bar{p}}{2}(1 - \frac{(\bar{p})^2}{2})$$

which holds for $\lambda \geq \frac{4}{7}$. Since $\lambda \in (\frac{4}{7}, \frac{3}{4})$, this inequality for all λ under consideration. Therefore, the expected utility under the separating equilibrium is strictly greater than that of semi-pooling equilibrium.

C Assumptions on Distributions

This section relaxes the assumption on the distribution of policy preferences being $U[-1,1]$. Specifically, it checks bounds on λ and existence of equilibria when assuming a general uniform distribution and a normal distribution with mean 0.

General Uniform Distribution.
First, consider a more general form of uniform distribution: $U[-D,D]$. For the separating equilibrium, the boundary condition on λ remains the same:

$$\underline{\lambda} \equiv \frac{1}{1+\sigma^2} = \frac{1}{1+\frac{D^2}{3}} = \frac{3}{3+D^2}$$

For the semi-pooling equilibrium, solving for $\bar{p}$ using the indifference condition and imposing the requirement that $|\bar{p}| > 0$ yields its boundary condition on λ:

$$\bar{\lambda} \equiv \frac{4}{4+3D^2}$$

Therefore, $\bar{\lambda} \leq \underline{\lambda}$, meaning the equilibrium results derived above applies for this general case.

Normal Distribution with Mean 0.
This part carries out the same exercise but assuming a normal distribution $N(0,\sigma^2)$. For

the separating equilibrium, the boundary condition on λ remains the same.

Now consider the separating equilibrium. To reduce notations, define $\phi(\cdot)$ and $\Phi(\cdot)$ as the pdf and cdf of a standard normal distribution, respectively. In addition, define $P_1 = \frac{-\bar{p}}{\sigma}$ and $P_2 = \frac{\bar{p}}{\sigma}$. Then, the mean and the variance of the normal distribution truncated at $-\bar{p}$ and $\bar{p}$ are:

$$\text{Mean} = 0 + \frac{\phi(P_1) - \phi(P_2)}{\Phi(P_2) - \Phi(P_1)}\sigma$$

$$\text{Var} = \sigma^2[1 + \frac{P_1\phi(P_1) - P_2\phi(P_2)}{\Phi(P_2) - \Phi(P_1)} - (\frac{\phi(P_1) - \phi(P_2)}{\Phi(P_2) - \Phi(P_1)})^2]$$

Then, the voter's expected utility from each of candidate B's action becomes:

$$E[u_v|(\varnothing, \{O\})] = -\lambda((\frac{\phi(P_1) - \phi(P_2)}{\Phi(P_2) - \Phi(P_1)}\sigma)^2 + \sigma^2[1 + \frac{P_1\phi(P_1) - P_2\phi(P_2)}{\Phi(P_2) - \Phi(P_1)} - (\frac{\phi(P_1) - \phi(P_2)}{\Phi(P_2) - \Phi(P_1)})^2])$$

$$E[u_v|(p_A, \varnothing)] = -\lambda p_A^2 - \frac{(1-\lambda)}{2}$$

As a reminder, $\bar{p}$ solves:

$$E[u_v|(\varnothing, \{O\})] = E[u_v|(\bar{p}, \varnothing)] \Leftrightarrow$$

$$-\lambda((\frac{\phi(P_1) - \phi(P_2)}{\Phi(P_2) - \Phi(P_1)}\sigma)^2 + \sigma^2[1 + \frac{P_1\phi(P_1) - P_2\phi(P_2)}{\Phi(P_2) - \Phi(P_1)} - (\frac{\phi(P_1) - \phi(P_2)}{\Phi(P_2) - \Phi(P_1)})^2]) = -\lambda(\bar{p})^2 - \frac{(1-\lambda)}{2}$$

$$-\lambda[\sigma^2(1 - \frac{\sqrt{\frac{2}{\pi}}\exp(-\frac{(\bar{p})^2}{2})\bar{p}}{\sigma \text{erf}(\frac{\bar{p}}{\sqrt{2}})})] = -\lambda(\bar{p})^2 - \frac{1-\lambda}{2}$$

Solving the equality above for $\bar{p}$,

$$\bar{p} \equiv \text{[solution here]}$$

D Extension 1 Proofs: Correlated Types

This section provides formal proofs of the extended model where an individual's moral and policy types are correlated.

D.1 Cases 1 and 2: $\lambda = 0$ and $\lambda = 1$

Proofs are omitted, as the same logic from the baseline model applies; the voter expecting the candidates to always campaign on moral messages or policy platforms can form posteriors that prevent candidates from deviating to other strategies (e.g., they always campaign on moral messages when $\lambda = 0$). With the voter only caring about one aspect of the candidate, the correlated types does not affect the equilibrium in these settings.

D.2 Case 3: $\lambda \in (0, 1)$

Equilibrium 1 – Separating
For the separating equilibrium, the proof remains the same for the simple reason that candidates are predicted to pool in terms of their strategies by moral types. This implies that correlation in types does not deliver additional information about the candidate. Therefore, the same proof from the baseline applies.

Equilibrium 2 – Semi-Pooling
Now conjecture an equilibrium where even the morally-aligned candidates with policy preferences close enough to that of the voter are expected to campaign on policies. Denote the cut-points where the A-type candidates are indifferent between campaigning on policies and moral messages as $\hat{p}_1$ and $\hat{p}_2$ for the left and the right side of the policy spectrum, respectively. D types are conjectured to campaign solely on policies. Then, the voter's posteriors are as follows:

$$\mu[t_j \neq t_v|(p_j, \varnothing), p_j \in [\hat{p}_1, 0)]] = \pi$$

$$\mu[t_j \neq t_v|(p_j, \varnothing), p_j \in (0, \hat{p}_2)]] = 1 - \pi$$

$$\phi[(\varnothing, O)] = \pi U[\hat{p}_2, 1] + (1 - \pi)U[-1, \hat{p}_1]$$

$$\phi[(\varnothing, U)] = 1$$

Based on these posteriors, consider the voter's expected utility from each of the possible action by A types:

$$E[u_v|(\varnothing, O)] = -\lambda[\pi \cdot ((\frac{\hat{p}_2 + 1}{2})^2 + \frac{(1 - \hat{p}_2)^2}{12}) + (1 - \pi) \cdot ((\frac{\hat{p}_1 - 1}{2})^2 + \frac{(1 + \hat{p}_1)^2}{12})]$$

$$E[u_v|(\hat{p}_1, \varnothing)] = -\lambda(\hat{p}_1)^2 - (1 - \lambda)\pi$$

$$E[u_v|(\hat{p}_2, \varnothing)] = -\lambda(\hat{p}_2)^2 - (1 - \lambda)(1 - \pi)$$

$\hat{p}_1$ solves the equality between the first two expected utilities, and $\hat{p}_2$ solves the equality between the first and the third expected utilities. The boundary condition on λ can be derived from the analytical solutions of $\hat{p}_1 < 0$ and $\hat{p}_2 > 0$.[37] Figure 2 provides the numerical solutions of these cut-points for different values of π, and the characterization of the equilibrium is provided in the main text.

Existence of Other Equilibria
Just as in the baseline case, there exists equilibria where voter's (relatively) extreme posteriors can sustain the one-sided equilibrium where both types of candidates only campaign on moral or policy platforms. Proofs remain unchanged.

E Extension 2 Proofs: Introducing the Zero-Cost Types

This section provides the formal proofs for the results of the extended model that introduces the zero-cost type candidates who do not incur any cost from announcing policy or moral

[37] The analytical solution is omitted to conserve space.

types different from their true types.

E.1 Case 1: $\lambda = 0$

There are zero-cost type candidates who do not incur any cost for campaigning on policies or moral value that do not correspond with their actual types. As a reminder, the probability of being a zero-cost type candidate is $Pr(Z=1) = q \in (0,1)$, and the voter's posterior on the probability of a candidate j being a zero-cost type upon observing his campaign message is $\eta[Z|(x_j, m_j)]$.

Just as in result 1, the voter does not care about the policy alignment at all. Conjecture an equilibrium where the infinite-cost and advantaged and zero-cost types always campaign on aligned moral message, while infinite-cost and disadvantaged types are indifferent between the either action. As the infinite-cost types cannot lie, they do not have incentives to deviate for the same reason as result 1.

Now consider zero-cost types' incentive to deviate by sending a policy message instead. First, note the voter's posterior on candidate's moral alignment upon observing aligned moral message (i.e., $(\varnothing, O)$):

$$\mu[t_j \neq t_v|(\varnothing, O)] = 1 - Pr(t_j = O|(\varnothing, O)) = 1 - \frac{\frac{1}{2}(q + (1-q))}{\frac{1}{2}1 + \frac{1}{2}q} = 1 - \frac{1}{1+q} = \frac{q}{1+q}$$

Based on the initial conjecture, voter's posterior on moral misalignment upon observing a policy platform announcement (i.e., $(p_j, \varnothing)$) remains the same as result 1:

$$\mu[t_j \neq t_v|(p_j, \varnothing)] = 1$$

Then, the voter's expected utilities from each action are:

$$E[u_v|(\varnothing, O)] = -\frac{q}{1+q}$$

$$E[u_v|(p_j, \varnothing)] = -1$$

Since $q \in (0,1)$, the voter strictly prefers a candidate who sends the aligned moral message. Therefore, the zero-cost type would not deviate by campaigning on policy platforms. Then, the characterization of the equilibrium is as follows:

- $\gamma^*_{A,I} = 0 : (x^*_{A,I}, m^*_{A,I}) = (\varnothing, O)\ \forall p_{A,I} \in P$; morally advantaged infinite-cost candidates (A, I) always campaign on moral messages.
- $\gamma^*_{D,I} = \begin{cases} 1 : (x^*_{D,I}, m^*_{D,I}) = (p_{D,I}, \varnothing), & \text{with probability } \frac{1}{2} \\ 0 : (x^*_{D,I}, m^*_{D,I}) = (\varnothing, U), & \text{with probability } \frac{1}{2} \end{cases} \forall p_{D,I} \in P$; morally disadvantaged infinite-cost candidates (D) are indifferent between campaigning on policy or moral messages.[38]

[38] Just as in *Result 1*, any strategy constitutes an equilibrium behavior for D, I candidates because no

- $\gamma^*_{A,Z} = \gamma^*_{D,Z} = 0 : (x^*_{A,Z}, m^*_{A,Z}) = (x^*_{D,Z}, m^*_{D,Z}) = (\varnothing, O)\ \forall p_{A,Z}, p_{D,Z} \in P$
- $r^*_v = \begin{cases} 1, & \text{if } (x_1, m_1) = (\varnothing, O) \neq (x_2, m_2) \\ \frac{1}{2}, & \text{if } \begin{cases} (x_1, m_1) = (x_2, m_2) \\ (x_1, m_1) \neq (x_2, m_2) \wedge m_1 = m_2 = \{\varnothing\} \end{cases} \\ 0, & \text{otherwise.} \end{cases}$
- Voter posterior beliefs:
  - $\mu^*_v[t_j \neq t_v | (p_j, \varnothing)] = 1$; voter believes candidate j to be morally misaligned if she observes policy platform instead of moral message.
  - $\mu^*_v[t_j \neq t_v | (\varnothing, O)] = \frac{q}{1+q}$: voter's belief on candidate's moral type being misaligned with hers upon observing $((\varnothing, O))$

E.2 Case 2: $\lambda = 1$

Now the voter only cares about the policy alignment. Conjecture an equilibrium where zero-cost types mix over some interval $[-\hat{p}, \hat{p}]$, where $\hat{p} \leq 1$, while the infinite-cost types adopt the pure strategy of campaigning truthfully on policy platforms just as in result 2.

For zero-cost types to be indifferent between campaigning on a particular policy platform within the interval, the voter's expected utility from such policy announcements should be constant. Define zero-cost candidate's mixing probability as $f(p'_z)$, where $p'_z \in [-\hat{p}, \hat{p}]$. Formally, this indifference can be represented as:

$$E[u_v | (p, \varnothing), |p| \in [0, \hat{p}]] = -(\frac{qf(p)}{qf(p) + \frac{(1-q)}{2}} \cdot Var(p) + \frac{\frac{(1-q)}{2}}{qf(p) + \frac{(1-q)}{2}} \cdot p^2) =$$

$$-\frac{qf(0)}{qf(0) + \frac{(1-q)}{2}} \cdot Var(p) = E[u_v | (0, \varnothing)]$$

That means voter's expected utility from seeing policy 0 and other policy in the given interval should be the same. Denote the LHS constant values as B. Using the fact that $Var(p) = \frac{1}{3}$ and solving for $f(p)$ yields:

$$f(p) \equiv \frac{(1-q)(B - p^2)}{q(\frac{1}{3} - B)}$$

Note that $\int_0^{\hat{p}} f(\tilde{p}) d\tilde{p} = \frac{1}{2}$, since all zero-cost types are assumed to mix over the given interval. Also, $f(\hat{p}) = 0$; the zero-cost types' mixing probabilities are decreasing in p, as they would find it a best response to campaign on policies closer to the voter's ideal point. Substituting $\hat{p}$ for p in $f(p)$ yields: $\hat{p}^2 = B$.

strategy is weakly dominated for them; they are expected to lose to the morally-aligned and zero-cost type candidates regardless of their actions. Since the voter does not care about policy at all, the "goodness" of one's policy type does not affect the relative appeal of either action.

The expression for $\hat{p}^2$ and the integral value constitute a system of equations that can be used to solve for $f(0)$ and $\hat{p}$. Taking the integral, plugging in the actual values for B and substituting in $\hat{p}$ yields:

$$\hat{p}(f(0) - \frac{qf(0) + \frac{(1-q)}{2}}{q} \cdot \frac{qf(0)}{3(qf(0) + \frac{(1-q)}{2})}) = \frac{1}{2} \Leftrightarrow \hat{p} = \frac{3}{4f(0)}$$

Substituting the expression back to $\hat{p}^2 = B$ yields:

$$0 < \hat{p} < \frac{1}{\sqrt{3}}, q \equiv \frac{4\hat{p}^3}{4\hat{p}^3 - 3\hat{p}^2 + 1}$$

As $q \to 1$, $\hat{p} \to \frac{1}{\sqrt{3}}$. Also note that for any policy announced in the given interval, the voter's expected utility is $-\hat{p}^2$. In other words, whenever she observes policy $|\hat{p} + \epsilon|, \epsilon > 0$, she prefers to elect the potential liar over an honest but extreme candidate. This prevents the zero-cost types from deviating. The voter can form the following posterior (from result 2) to prevent candidates from campaiging on moral messages:

$$\phi[|p_j||(x_j, m_j) = (\varnothing, t_j)] = 1$$

The characterization of the equilibrium is reported in the text.

Claim: There does not exist a pure strategy equilibrium where zero-cost types pool at any particular point.

Suppose not, and conjecture an equilibrium where zero-cost types pool at any point p_z. No matter where they pool, any zero-cost type can profitably deviate by campaigning on policy just left or right of the mass point (p') to make himself appear as an infinite-cost type to the voter, and the voter would prefer such a candidate over those who report the conjectured pool point. Therefore, such a pure strategy equilibrium where zero-cost types pool at any given point is not possible. □

E.3 Case 3: $\lambda \in (0, 1)$

Note that voter now cares about both policy and moral alignment.

Equilibrium 1 – Separating: Just as in result 3 equilibrium 1, conjecture an equilibrium where morally-advantaged types always campaign on moral messages and disadvantaged types campaign on policy platforms. The zero-cost types, whether they are advantaged or disadvantaged, mimic the advantaged types. Then, the voter's posterior on the candidate's moral misalignment upon observing $(\varnothing, O)$ becomes:

$$\mu[t_j \neq t_v|(\varnothing, O)] = 1 - Pr(t_j = O|(\varnothing, O)) = 1 - \frac{\frac{1}{2}(q + (1-q))}{\frac{1}{2}1 + \frac{1}{2}q} = 1 - \frac{1}{1+q} = \frac{q}{1+q}$$

For the infinite-cost morally aligned and zero-cost types to not deviate to sending policy message, consider the best possible policy type; if the expected utility from campaigning on aligned moral message is greater than that of the best possible policy message, neither would deviate.

$$E[u_v|(0, \varnothing)] = -(1-\lambda)$$
$$E[u_v|(\varnothing, O)] = -\lambda\frac{1}{3} - (1-\lambda)(\frac{q}{1+q})$$

Setting the inequality for the latter to be greater yields the boundary condition on λ that can sustain this equilibrium:

$$\lambda < \frac{3}{1+q+3} \equiv \underline{\lambda}$$

This suggests that the boundary is decreasing in q; as the probability of being a zero-cost candidate increases, the values of λ that can sustain the equilibrium decreases. The characterization of the equilibrium is reported in the text.

Equilibrium 2 – Semi-Pooling:
Now conjecture an equilibrium similar to that of result 3. By the same logic from the claim from result 5, there cannot be an equilibrium where zero-cost types pool at any particular point when campaigning on policies. That means, if they are to campaign on policies at all, they must be mixing over a certain interval. This necessarily implies that the voter's expected utility from observing any "good" policy should be constant. For the infinite-cost types, conjecture that the morally-advantaged type adopts a cut-point strategy where below some threshold $\hat{\hat{p}}$, they send policy messages, and moral, otherwise. Suppose the morally-disadvantaged infinite-cost types always campaign on policies, just as in the previous construction.

Based on this initial conjecture, then, the A, I types with a policy preference exactly at $\hat{\hat{p}}$ must be indifferent between campaigning on policy and moral messages. However, note that the zero-cost types (Z) are conjectured to mix over the interval and the voter's expected utility is constant. That means, for all $p_{A,I}$ s.t. $|p_{A,I}| \in [0, \hat{\hat{p}}]$, the A, I types must be indifferent between campaigning on policy and moral messages. This, then, further implies that Z types must be also indifferent between the two actions.

Define A, I types' mixing probability of selecting policy as α and Z types' mixing probabilities as β and $f(p)$ for selecting policy and mixing over policies, respectively. Then, voters' posteriors upon observing each set of campaign actions are as follows:

- $\phi_v^{[}(\varnothing, U)] = 1$; voter believes candidate j's policy type to be the worst possible type upon observing that he reveals his moral misalignment.
- $\phi_v[(\varnothing, O)] = \begin{cases} U[-1, 1] & \text{with prob. } q \\ U[-1, -\hat{\hat{p}}] + U[\hat{\hat{p}}, 1] & \text{with prob. } 1-q \end{cases}$
- $\phi_v[(p_j, \varnothing), |p_j| \in [0, \hat{\hat{p}}]] = \begin{cases} U[-1, 1] & \text{with prob. } q \\ p_j & \text{with prob. } 1-q \end{cases}$

- $\eta_v[Z|(\varnothing, O)] = \frac{q(1-\beta)}{q(1-\beta)+\frac{1-q}{4}(\hat{p}(1-\alpha)+1-\hat{p})}$

- $\eta_v[Z|(p_j, \varnothing), |p_j| \in [0, \hat{\hat{p}}]] = \frac{q\beta f(p)}{q\beta f(p)+\frac{1-q}{4}(1+\alpha)}$

- $\mu_v[t_j \neq t_v|(\varnothing, O)] = \frac{q(1-\beta)}{2q(1-\beta)+\frac{(1-q)}{2}(\hat{p}(1-\alpha)+1-\hat{p})}$; voter's posterior on candidate's moral type being aligned upon observing that the candidate campaigns on aligned moral messages.

- $\mu_v[t_j \neq t_v|x_j \in [0, \hat{\hat{p}}]] = \frac{q\beta f(p)+\frac{(1-q)}{2}}{2q\beta f(p)+\frac{(1-q)}{2}(1+\alpha)}$; voter's posterior on candidate's moral type upon observing that the candidate campaigns on policy platforms (in absolute terms) that fall in the interval $[0, \hat{\hat{p}}]$

- $\mu_v[t_j \neq t_v|x_j \in (\hat{\hat{p}}, 1]] = 1$; voter believes candidate to be morally misaligned upon observing that the candidate campaigns on policy platforms (in absolute terms) that fall in the interval $(\hat{\hat{p}}, 1]$.

Based on these posteriors, voters' expected utilities upon observing $(p, \varnothing); |p| \in [0, \hat{\hat{p}}]$, $(0, \varnothing)$, and $(\varnothing, O)$ are as follows:

$$E[u_v|(p, \varnothing); |p| \in [0, \hat{\hat{p}}]] = -\lambda[\frac{q\beta f(p)}{q\beta f(p) + \frac{(1-q)}{2}\frac{(1+\alpha)}{2}} \cdot Var(p) + \frac{\frac{(1-q)}{2}\frac{(1+\alpha)}{2}}{q\beta f(p) + \frac{(1-q)}{2}\frac{(1+\alpha)}{2}} \cdot p^2]$$

$$-(1-\lambda)\frac{q\beta f(p) + \frac{(1-q)}{2}}{q\beta f(p) + \frac{(1-q)}{2}\alpha + q\beta f(p) + \frac{(1-q)}{2}}$$

$$E[u_v|(0, \varnothing)] = -\lambda[\frac{q\beta f(0)}{q\beta f(0) + \frac{(1-q)}{2}\frac{(1+\alpha)}{2}} \cdot Var(p)] - (1-\lambda)\frac{q\beta f(0) + \frac{(1-q)}{2}}{q\beta f(0) + \frac{(1-q)}{2}\alpha + q\beta f(0) + \frac{(1-q)}{2}}$$

$$E[u_v|(\varnothing, O)] = -\lambda[\frac{q(1-\beta)}{q(1-\beta) + \frac{(1-q)}{2}\frac{1}{2}(\hat{\hat{p}}(1-\alpha) + 1 - \hat{\hat{p}})} \cdot Var(p) +$$

$$\frac{\frac{(1-q)}{2}\frac{1}{2}(\hat{\hat{p}}(1-\alpha) + 1 - \hat{\hat{p}})}{q(1-\beta) + \frac{(1-q)}{2}\frac{1-\alpha}{2}} \cdot ((\frac{\hat{\hat{p}} + 1}{2})^2 + \frac{(1-\hat{\hat{p}})^2}{12})] - (1-\lambda)[\frac{q(1-\beta))}{2q(1-\beta) + \frac{(1-q)}{2}(\hat{\hat{p}}(1-\alpha) + 1 - \hat{\hat{p}})}]$$

For Z types to mix over the interval $[-\hat{\hat{p}}, \hat{\hat{p}}]$, it follows that the first two expected utilities should be the same:

$$E[u_v|(p, \varnothing); |p| \in [0, \hat{\hat{p}}]] = E[u_v|(0, \varnothing)] = \tilde{B}$$

Solving the expression for the mixing probability $f(p)$ yields:

$$f(p) = \frac{\frac{(1-q)}{2}((1+\alpha)\tilde{B} + (1 - \lambda + \lambda p^2(1+\alpha)))}{-q\beta(1 - \frac{\lambda}{3} - 2\tilde{B})}$$

Using the assumption that $f(\hat{\hat{p}}) = 0$ (i.e., Z types assign zero probability to the boundary or the worst policy within the interval), the numerator portion yields the expression for $\hat{\hat{p}}$:

$$(1+\alpha)\tilde{B} + (1 - \lambda + \lambda\hat{\hat{p}}^2(1+\alpha) = 0 \Leftrightarrow \hat{\hat{p}}^2 = \frac{1}{\lambda}(-\tilde{B} - \frac{(1-\lambda)}{(1+\alpha)}) \rightarrow \text{Equation 1.}$$

Now utilizing the fact that $\int_0^{\hat{\hat{p}}} f(p)dp = \frac{1}{2}$,

$$\int_0^{\hat{\hat{p}}} f(p)dp = \frac{\frac{(1-q)}{2}}{-q\beta(1-\frac{\lambda}{3}-2\tilde{B})}\hat{\hat{p}}[(1+\alpha)\tilde{B} + (1-\lambda) + \frac{\lambda(1+\alpha)}{3}\hat{\hat{p}}^2] = \frac{1}{2} \rightarrow \text{Equation 2.}$$

Returning to the third expected utility, note that α – the mixing probability for the A, I types – solves:

$$E[u_v|(\varnothing, O)] = \tilde{B} \rightarrow \text{Equation 3.}$$

And β – the mixing probability for the Z types – solves:

$$E[u_v|(p, \varnothing); |p| \in [0, \hat{\hat{p}}]] = E[u_v|(\varnothing, O)] \rightarrow \text{Equation 4.}$$

Numerically solving these four simultaneous equations for α, β, $\hat{\hat{p}}$, and $f(0)$ yield figures reported in the text. The characterization of this equilibrium is reported in the main text.

Existence of other equilibria

Now we check the existence of other equilibria. Just as in the baseline case, there exists at least two other equilibria that share the same characterizations as *case 1* and *case 2* from above.

First, conjecture an equilibrium similar to that of *case 1* while still assuming $\lambda \in (0, 1)$; all types are conjectured to campaign on moral messages regardless of their policy preferences. Then, the voter would form the following posterior upon observing policy platform announcements from candidate j:

$$\mu[t_j \neq t_v|(p_j, \varnothing)] = 1$$

Then, following the logic from case 1, the Z and A, I types would consider it a best response to always campaign on moral messages, while the D, I types are indifferent between either action, as their misalignment is bound to be discovered. As a result, for all values of λ, the same exact characterization as *case 1* applies. □

Now conjecture an equilibrium similar to that of *result 5* while still assuming $\lambda \in (0, 1)$; all types are conjectured to campaign on policy platforms regardless of their moral alignment. Suppose Z types adopt the mixing strategy similar to that of result 5 and denote the cutpoint as $\tilde{p}$. Based on this initial conjecture, define voters' posteriors on the off-path behavior (e.g., $(\varnothing, t_j)$) as follows:

$$\mu[t_j \neq t_v|(\varnothing, t_j)] = 1$$

$$\phi[(\varnothing, t_j)] = 1$$

That is, whenever the voter observes a moral message, she assumes the candidate to be misaligned with probability 1 and to hold the most extreme policy preference. Then,

$$E[u_v|(\varnothing, O)] = -\lambda - (1-\lambda)$$

As this constitutes the lowest possible expected utility, deviating to campaigning on moral messages is dominated for all types of candidates. With the only change from result 5 being $\lambda \in (0,1)$, the same equilibrium characterization applies. □

F Extension 3 Proof: Correlated Types and Cheap Talkers

Conjecture a semi-pooling equilibrium where candidates are expected behave in a manner similar to those in *Results 6*. Based on such initial conjectures, voters' posteriors are as reported in the main text. The voter's expected utilities from each action is as follows:

$$E[u_v|(\varnothing, O)] =$$

$$-\lambda[\frac{q(1-\beta)}{q(1-\beta)+\frac{1-q}{2}(\frac{\hat{p}_2-\hat{p}_1}{2}(1-\alpha)+\frac{\hat{p}_1+1+1-\hat{p}_2}{2})}\cdot Var(p)+\frac{\frac{1-q}{2}(\frac{\hat{p}_2-\hat{p}_1}{2}(1-\alpha)+\frac{\hat{p}_1+1+1-\hat{p}_2}{2})}{q(1-\beta)+\frac{1-q}{2}(\frac{\hat{p}_2-\hat{p}_1}{2}(1-\alpha)+\frac{\hat{p}_1+1+1-\hat{p}_2}{2})}\cdot$$

$$\{\frac{\hat{p}_2-\hat{p}_1}{2}((\frac{\hat{p}_2+\hat{p}_1}{2})^2+\frac{(\hat{p}_2-\hat{p}_1)^2}{12})+\frac{\hat{p}_1+1}{2}((\frac{\hat{p}_1-1}{2})^2+\frac{(\hat{p}_1+1)^2}{12})+\frac{1-\hat{p}_2}{2}((\frac{\hat{p}_2+1}{2})^2+\frac{(1-\hat{p}_2)^2}{12})\}]+$$

$$(1-\lambda)[\frac{q(1-\beta)}{2q(1-\beta)+(1-q)(\frac{\hat{p}_1+1+1-\hat{p}_2}{2})}]$$

Upon observing $(p_j, \varnothing), p_j \in [\hat{p}_1, 0)$:

$$E[u_v|(p_j, \varnothing), p_j \in [\hat{p}_1, 0)] =$$

$$-\lambda[\frac{q\beta f(p)}{q\beta f(p)+\frac{1-q}{2}(\pi+(1-\pi)\alpha)}\cdot Var(p)+\frac{\frac{1-q}{2}(\pi+(1-\pi)\alpha)}{q\beta f(p)+\frac{1-q}{2}(\pi+(1-\pi)\alpha)}\cdot(p_j)^2]-$$

$$(1-\lambda)[\frac{q\beta f(p)+(1-q)\pi}{2q\beta f(p)+(1-q)((1-\pi)\alpha+\pi)}]$$

Upon observing $(p_j, \varnothing), p_j \in (0, \hat{p}_2]$:

$$E[u_v|(p_j, \varnothing), p_j \in (0, \hat{p}_2]] =$$

$$-\lambda[\frac{q\beta f(p)}{q\beta f(p)+\frac{1-q}{2}(\pi\alpha+(1-\pi)}\cdot Var(p)+\frac{\frac{1-q}{2}(\pi\alpha+(1-\pi))}{q\beta f(p)+\frac{1-q}{2}(\pi\alpha+(1-\pi))}\cdot(p_j)^2]-$$

$$(1-\lambda)[\frac{q\beta f(p)+(1-q)(1-\pi)}{2q\beta f(p)+(1-q)((1-\pi)+\pi\alpha)}]$$

Conjecture the constant value $\hat{B}$ that the voter receives upon observing $(0, \varnothing)$ to be:

$$E[u_v|(0,\varnothing)] = -\lambda[\frac{q\beta f(0)}{qbetaf(0) + \frac{(1-q)}{2}(1+\alpha)} \cdot Var(p)] - (1-\lambda)[\frac{q\beta f(0) + (1-q)}{2q\beta f(0) + (1-q)(1+\alpha)}] \equiv \hat{B}$$

Following the same algorithm as the *Result 6*, we can construct the system of equations as follows:

1. $E[u_v|(p_j, \varnothing), p_j \in [\hat{p}_1, 0)] = \hat{B} \Rightarrow \hat{p}_1$ solves:

$$\hat{B}(\pi + (1-\pi)\alpha) - (-\pi + \lambda(\pi - (\hat{p}_1)^2(\pi + (1-\pi)\alpha))) = 0$$

2. $E[u_v|(p_j, \varnothing), p_j \in (0, \hat{p}_2]] = \hat{B} \Rightarrow \hat{p}_2$ solves:

$$\hat{B}(\pi\alpha + (1-\pi)) - (-(1-\pi) + \lambda((1-\pi) - (\hat{p}_2)^2(\pi\alpha + (1-\pi)))) = 0$$

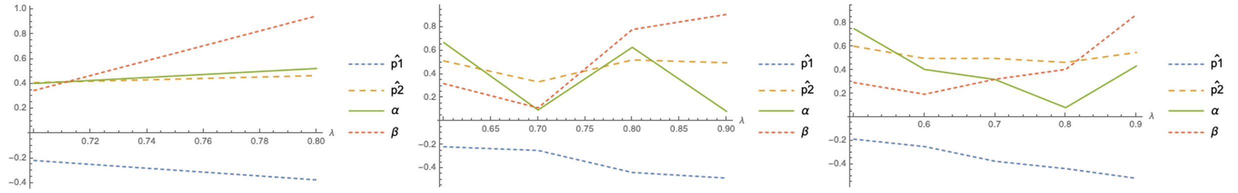
3. Based on the assumption on the Z types' mixing probability that $\int f(p) = 1$,

$$\int_{\hat{p}_1}^{0} f(p)dp + \int_{0}^{\hat{p}_2} f(p)dp = 1$$

4. $E[u_v|(\varnothing, O)] = \hat{B}$

5. $E[u_v|(\varnothing, O)] = \pi E[u_v|(p_j, \varnothing), p_j \in (0, \hat{p}_2]] + (1-\pi)E[u_v|(p_j, \varnothing), p_j \in [\hat{p}_1, 0)]$

With these five equations and five unknown parameters ($\alpha, \beta, \hat{p}_1, \hat{p}_2$, and $f(0)$), solving them simultaneously while assuming specific values for exogenous parameters λ, π, and q return numerical solutions plotted in figure 6 below.[39]



(a) $q = \frac{1}{4}$

(b) $q = \frac{1}{2}$

(c) $q = \frac{3}{4}$

Figure 6: Numerical Solutions of $\hat{p}_1, \hat{p}_2, \alpha, \beta$ for Various q, assuming $\pi = \frac{2}{3}$

[39] Note that these are relatively "low" resolution results that only consider 10 different values of λ (e.g., $\frac{1}{10}$ to $\frac{10}{10}$. The main text includes a figure with a higher resolution that considers 100 different values of λ.