

A THEORY OF DECISIVE LEADERSHIP

B. DOUGLAS BERNHEIM AND AARON L. BODOH-CREED

ABSTRACT. We present a theory that rationalizes voters' preferences for decisive leaders. Greater decisiveness entails an inclination to reach decisions more quickly conditional on fixed information. Although speed can be good or bad, agency problems between voters and politicians create preferences among voters for leaders who perceive high costs of delay and have little uncertainty about how to weigh different aspects of the decision problem, and hence who make decisions more rapidly than typical voters. Officials who aspire to higher office therefore signal decisiveness by accelerating decisions. In elections, candidates with reputations for greater decisiveness prevail despite making smaller compromises, and therefore earn larger rents from office holding.

“When we evaluate candidates, decisiveness is one of the words that bubbles up most frequently... [O]ur new leader must render decisions on our behalf that are swift, strong, and sure... We favor ‘Damn the torpedoes, full speed ahead,’ and reject ‘Let me mull it over and I’ll get back to you after I weigh the options’... The very appearance of hesitation can sink a presidency... And it also can sink a presidential campaign.” Ted Anthony, “The cult of decisiveness in U.S. politics,” NBCNEWS.com, April 4, 2008.¹

1. INTRODUCTION

During elections, voters choose to support candidates partly based on stated positions concerning well-established issues. However, the voters also know new issues will arise after the winner of the election takes office, some of which will be entirely unanticipated. Consequently, voter support depends in part on perceptions of characteristics that bear on a candidate’s ability to handle emerging situations and crises. In this paper, we are concerned with one such characteristic, *decisiveness*.

The literature on leadership defines decisiveness as the ability to come to a timely decision despite uncertainty (Simpson, French, and Harvey [30], Simon [29], Williams et al. [31]). According to this literature, the public tends to perceive leaders as effective when they exhibit a “bias for action” rather than “paralysis by analysis” (Kelman et al. [18]).

¹<http://www.nbcnews.com/id/23884031/ns/politics/t/cult-decisiveness-us-politics/#.Wt5cucGouUk>, accessed April 25, 2018.

Indeed, Simon [29] goes so far to claim that, in contexts where actions must be taken quickly, “[behavioral ambivalence is unacceptable. . . Leaders who vacillate will not retain their positions.” As a result, “. . . dominant societal conditions militate against restraint and inaction on the part of leaders, however reflective their intent” (Simpson et al. [30]).²

Existing research on decisiveness falls almost entirely outside of economics. Consequently, it does not generally employ the empirical methods that economists favor. Instead, it relies heavily on descriptive analysis and case studies. Still, the observations made in this literature are provocative, and at minimum raise the possibility that economists have to date overlooked a potentially interesting and important topic. For example, dominant perspectives expressed in popular media purportedly applaud decisive leadership on the grounds that it promotes timely and consistent action (Williams et al. [31]).³ The political costs of indecisiveness are thought to be especially high during crises because “[speed matters, and time is a leader’s enemy. . .” (Garcia [14]; see also Yukl [32], Holsti [15], and Pillai and Meindl [23]). Opinion surveys show that voters place substantial weight on perceived decisiveness,⁴ and statistical analysis confirms that this perception predicts voting behavior (Williams et al. [31]).⁵ As a result, labeling a leader “indecisive” is widely construed as an indictment of his or her suitability for office.⁶

An interesting feature of the literature is its emphasis on the *speed* of decision making. It would be problematic to assume that voters value speed simply for its own sake. As a general matter, greater speed in the face of uncertainty can be either good or bad. In principle, effective leadership should require “negative capability,” defined as “the capacity to sustain reflective inaction,” as well as “positive capability,” the capacity for action (Simpson et al. [30]). A good theory of the preference for decisive leaders must explain

²These observations encompass business enterprises, the public sector, and even voluntary organizations: “Although the language used in the different sectors varies, the imperative is similar” (Simpson et al., 2002).

³Indeed, some studies have concluded that decisive leaders are more effective; see Deluga [12] and Kelman et al. [18].

⁴In one Gallup poll, respondents were asked to rate the importance of sixteen characteristics when evaluating presidential candidates. Being “a strong and decisive leader” was considered “essential” by 77% of respondents, highest of any characteristic, and “not that important” by 1% of respondents, lowest of any characteristic. (Jones [17]).

⁵Voter perception of politicians’ personal traits and their implications for electoral results has not attracted much attention within the political economy literature. Until recently, political scientists typically viewed voting based on a candidate’s personal characteristics as “irrational” (Shabad and Anderson [27]). However, recent studies have found that these characteristics significantly influence voters’ beliefs and behavior (Maurer et al. [21], Pillai and Williams [24], Pillai et al. [25], Shamir [28]). Most of the literature studies broadly defined leadership traits such as “charisma.” Rapoport, Metcalf, and Hartman [26] argue that voters make inferences about politicians’ policy preferences from their perceived traits, and vice versa.

⁶For example, Lindsey Graham criticized Barack Obama by calling him “a weak and indecisive president” (Washington Post, May 5, 2014).

why voters systematically applaud fast decision makers and deplore slow ones despite the pluses and minuses of speed.

Several theoretical accounts of decisiveness merit attention. One holds that decisive leaders make quick decisions because they have “sharp” worldviews, powerful moral compasses, and/or strong convictions. Under this view, indecisive leaders “waffle” or “flip-flop” because their criteria for aggregating costs and benefits are malleable and vary with the arrival of extraneous information. A second account attributes the greater speed of decisive decision makers to features of their preferences—for example, the weights they attach to opportunity costs or the steepness of their loss functions. A third account portrays decisive decision makers as enjoying informational advantages: either they are more knowledgeable at the outset, or they have superior technical (information-processing) skills.

In this paper, we adopt and articulate the “moral compass” and “preference” accounts of decisiveness. In our view, the third account, which equates decisiveness with technical aptitude, does not capture the term’s conventional meaning. As an example, consider the 2004 U.S. presidential race between George W. Bush and John Kerry, in which the issue of decisiveness took center stage. According to Democratic strategist James Carville, his Republican counterpart (Karl Rove) made the election “not about policies or positions or even about values or national security – he made it about decisiveness.”⁷ Few commentators described Bush as having an exceptional command of the issues; instead, his image as a “decisive” president emanated from the perception that he was a man of strong and set principles. Carville characterized the Republican message as, “You may not like what I stand for, but I stand for something.” To draw a sharp contrast, Senator Zell Miller of Georgia described John Kerry to the Republican National Convention as a “yes-no-maybe bowl of mush,” despite Kerry’s arguably superior technical expertise.⁸ According to one journalistic account, “Bush’s shoot-from-the-hip style... emphasizes decisiveness itself as a paramount trait, regardless of what the decisions might be... Bush’s stay-the-course message is presented as a position of strength. Any hint of indecisiveness – even the notion of considering multiple options before acting – is cast as weakness.”⁹ Commenting on the election, historian Bruce Schulman remarked, “It’s not so much decisiveness that is prized

⁷James Carville, “Karl Rove – A Brilliant (Ouch!) Political Strategist,” *Time* 165(16), p. 113.

⁸Jarrett Murphy, “Text of Zell Miller’s RNC Speech,” CBS News, September 1, 2004, <https://www.cbsnews.com/news/text-of-zell-millers-rnc-speech/>, accessed April 25, 2018.

⁹Ted Anthony, “The cult of decisiveness in U.S. politics,” NBCNEWS.com, April 4, 2008, <http://www.nbcnews.com/id/23884031/ns/politics/t/cult-decisiveness-us-politics/#.Wt5cucGouUk>, accessed April 25, 2018.

as it is commitment to principle... Flip-flop is bad... because it's seen as evidence of a lack of principle."¹⁰

This paper makes four main contributions. First, we formalize the notion of decisiveness and develop a theoretical framework suitable for studying it. Second, we show that a general electoral preference for decisive leaders, defined as a leader who is more willing than the typical citizen to reach decisions despite uncertainty, emerges naturally from the nature of the agency relationship between voters and the politician. Third, we point out that this preference creates an incentive for politicians to signal decisiveness by making highly visible decisions more rapidly, and we explore various implications of the signaling equilibrium, summarized below. Fourth, we demonstrate that perceived decisiveness enhances a leader's ability to impose his or her own agenda on voters. Greater political polarization magnifies this benefit, and therefore amplifies the incentive to appear decisive by rushing through deliberations.

With respect to our first contribution, a central feature of our framework is that political decision-makers can have either sharp or fuzzy world views. When an issue arises, the decision-maker takes time to learn about and evaluate its potentially multifaceted costs and benefits. Those with sharp views (e.g., a moral compass) know precisely how they feel about the relative importance of different costs and benefits; consequently, they reach decisions relatively quickly, and their choices are predictable. In contrast, those with fuzzy views struggle to evaluate trade-offs between different aspects of the decision. As a result, they reach decisions more slowly; their choices are not only less predictable, but also more susceptible to extraneous influences. We adopt a simple reduced-form representation of these differences, in which policy preferences consist of common and idiosyncratic components, both of which are initially random. Initial uncertainty about the idiosyncratic component is either high or low according to whether policy views are fuzzy or sharp. (Appendix B provides explicit microfoundations for this reduced-form.)

As an example, consider the issues facing U.S. officials during the financial crisis of 2008. As the prospect of widespread bank failures loomed, decision-makers had to weigh a variety of different concerns such as the risk of a widespread economic melt-down, the budgetary costs of rescuing large banks, and the moral hazard problems that would result from a general bailout. In such circumstances, a conservative with a sharp world view might decide against a bailout immediately based on free market principles, while a liberal with a sharp world view might quickly intervene to alleviate economic hardship. A politician

¹⁰Quoted in Ted Anthony, *op. cit.*

with a fuzzy outlook might agonize over the trade-offs, not only creating delay, but also making the outcome more difficult to predict *ex ante*.

Within our framework, decision speed is interpretable as an index of decisiveness because it captures the individual's willingness to take action at any given point in time conditional on a fixed rate of information acquisition. It varies from one person to another for two reasons. First, some people see the costs of deferring a decision as higher than others (aversion to delay), or equivalently have relatively flat loss functions. Second, some people have sharper views about values and tradeoffs than others. Decision makers with high delay aversion and sharp world views display a "bias for action," while those with low delay aversion and fuzzy world views exhibit "paralysis by analysis."¹¹

With respect to our second contribution, we show that, for each voter, the preferred politician perceives greater costs of delay than the voter and has a sharper world view, and as a result would make a given decision more rapidly than the voter. Because this statement holds for all voters, decisive candidates (defined as those who would make a given decision more rapidly than the median voter) have an electoral advantage over indecisive ones. Indeed, when politicians differ only with respect to decisiveness, there exist Condorcet-winning candidates, and they are all decisive decision makers.

Intuitively, the preference for decisive leaders is a natural outgrowth of the agency problem between voters and politicians. Voters do not benefit from any information the office holder uses to resolve a fuzzy world view. On the contrary, they face additional risk, since they are uncertain as to which preferences the indecisive politician will ultimately express, and hence which policy will be implemented. Consequently, each voter would ideally like to elect a politician with a sharp world view who shares the voter's degree of aversion to delay. Unfortunately, no politician has a perfectly sharp perspective. To reduce the risk arising from the resolution of fuzziness, the voter would like any office holder with a fuzzy view to make decisions more rapidly than the voter's ideal politician, who in turn decides more rapidly than the voter. Since all voters favor politicians who make decisions faster than themselves, a majority of the population would like to elect faster-than-average leaders who know their own minds and perceive greater-than-average costs of delay.

The preceding observation leads naturally to our third main point: politicians will attempt to signal decisiveness by accelerating their observable decisions in order to appeal to the electorate. We formalize this observation in a setting where heterogeneity is limited to delay aversion, focusing on the case in which the median voter thinks all politicians are

¹¹Decision speed could also vary if some people gather information more efficiently than others. That variation would not, however, reflect differences in decisiveness as the term is commonly understood. A more decisive individual decides more rapidly conditional on a fixed rate of information acquisition.

naturally inclined to make decisions too slowly.¹² We then consider implications for voter welfare, candidate self-selection, and transparency.

Because signaling increases decision speed, one might conjecture that it would benefit most voters. As we show, signaling is a rather poor solution for politicians' tendency to delay because its effect is smallest where the need for a corrective influence is greatest (i.e., when politicians are indecisive) and greatest where that need is smallest (i.e., when politicians are decisive). In the latter case, it can cause the politician to make decisions that are excessively hasty from every voter's perspective. Since voters favor politicians with high costs of delay, the voters' preferred politicians suffer higher costs and get lower equilibrium payoffs than politicians that are less averse to delay. Moreover, the fact that signaling costs fall primarily on decisive politicians has straightforward but significant implications for the distribution of politicians' characteristics in settings with endogenous candidacy: voters' least favorite potential candidates (those who are indecisive) receive the greatest rewards from holding office, and hence have the largest incentive to enter politics.

We define transparency as the ability of voters to observe the decision process underlying multiple choices made by the politician in a lower office prior to the current election. We show that greater transparency of decision-making leads politicians to signal through a pattern of consistent decisiveness rather than occasionally displaying extreme decisiveness. As a result, greater transparency reduces the signaling distortion and favorably resolves the welfare ambiguity associated with signaling. However, too much transparency is welfare-reducing for voters. From the perspective of institutional design, natural objectives therefore come into conflict: on the one hand, transparency promotes accountability; on the other hand, it may cause politicians to make decisions too slowly from the voters' perspective.

We make our fourth main point by extending the model to encompass the possibility that politicians with ex ante heterogeneous policy preferences can make ideological compromises by, for example, granting interest groups "seats at the table." In a Downsian-style electoral competition between two candidates whose policy inclinations are equally distant from those of the median voter, the more decisive candidate will prevail despite striking the more modest compromise. A greater ability to resist compromise translates into larger rents for the office holder. In effect, more decisive candidates impose their personal agendas on the electorate.

¹²We examine the case in which politicians signal the sharpness of their world view rather than an aversion to delay in Appendix C. Most of the results are similar, but there are some significant differences.

The remainder of this paper is organized as follows. Section 2 reviews related literature. Section 3 sets forth the basic model, and Section 4 uses it to rationalize voters' preference for decisiveness. Section 5 investigates the properties of equilibria wherein politicians signal decisiveness while in lower office. Section 6 extends our model to cover settings with ex ante heterogeneity among politicians' policy preferences, as well as mechanisms for ideological compromise. Section 7 provides some brief concluding remarks. All proofs appear in the appendix.

2. RELATED LITERATURE

Our model is situated within the political agency literature. Most of this literature focuses on the potential misalignment between the policies chosen by politicians and the preferences of the electorate. In contrast, we focus on misalignments between politicians and voters involving decision-making strategies.

Our paper touches on the growing decision-theoretic literature concerning indecision. Starting with Aumann [1], most authors have treated indecisiveness as incompleteness of the preference relation. From this perspective, a person is deemed indecisive between two options if he or she has no preference ranking between them. Recent papers in this literature elaborate on this basic premise. For example, Ok, Ortoleva, and Riella [22] explore the difference between indecisive tastes and indecisive beliefs. More recently some papers have characterized indecision as a preference for choice deferral (e.g., Kopylov [19]), but these models also involve incomplete preferences. Like Kopylov [19], we take the view that decisiveness and indecisiveness are inherently dynamic phenomena. In contrast to the decision-theoretic literature, we characterize people as less decisive if they either have greater uncertainty about their own (complete) preferences or perceive smaller costs of delay. In our model, those individuals take longer to make up their minds, even though everyone acquires and accurately processes information at the same rate.

Our analysis also relates to a branch of the literature originating with Conlisk [10] that explores settings in which agents must take time or pay costs to learn about their own preferences; see also Ergin and Sarver [11] for a more recent decision-theoretic treatment. Our contribution is embedding a related model of choice into a model of political agency and exploring the implications.

Our analysis also builds on themes found in the general literature concerning principal-agent problems. In the standard formulation, the principals and agents have diametrically opposed preferences concerning the agent's effort: the principal wants more, while the agent wants less. The principal designs compensation to counter the agent's proclivity

to slack off, and would plainly prefer an agent who is less averse to effort (equivalently, less inclined toward rent-seeking). In contrast, a key feature of our model is that the interests of the principals (voters) and agent concerning rent-seeking are aligned in some instances and misaligned in others. In our setting, voters perceive both costs and benefits to delay, and each one has in mind an ideal speed for the politician. In a sense, rent-seeking involves incremental delay to resolve the fuzziness of the politician’s world view. However, all delay—regardless of whether it is motivated by rent-seeking—informs the politician about common concerns. It follows that, from the point of view of the voter, only the total amount of delay matters. Thus, if the voter perceives a lower cost of delay than the politician, he will believe that the politician makes up his mind too fast absent rent-seeking motives, and he will be happier if rent-seeking slows the politician down sufficiently to achieve what the voter takes to be the ideal speed. An additional twist is that different voters have different preferences. Thus, some voters may wish the politician was more inclined to rent-see, while others may wish the politician was less inclined. Our results represent non-trivial extensions of the principal-agent literature in part because they hold despite these non-standard complications.

3. THE MODEL

3.1. Agents and preferences. We study environments with two types of agents, voters and politicians. Voters select among politicians in elections. Ultimately one politician takes office and selects a policy $p \in \mathbb{R}$. Agents’ preferences over policies depend on the state of the world, which is unknown at the outset. The politician spends time $\tau \geq 0$ gathering information and thinking about options before making a choice. Delay is costly, so agents also care about τ .

We can decompose the fully informed ideal policy for each voter into two components, one common (θ), the other idiosyncratic (x_i). The ideal point for agent i is $\theta + x_i$, which is reflected in the following utility function¹³

$$U_i(p, \tau) = -(p - x_i - \theta)^2 - c_i\tau,$$

where the quadratic term reflects policy preferences and the second term captures the cost of gathering information. Utility for the governing politician is given by the same expression with P substituting for i . The common component, θ , represents the ideal policy with full information according to average preferences. Returning to the example of the financial

¹³Bernhardt et al. [2] use a similar structure to model uncertainty about voters’ preferences, but their focus differs from ours.

crisis, the realization of θ might represent the magnitude of the ideal bailout, weighing the effects on economic activity, the government’s budget, and precedent as would the typical voter. All agents have a common prior belief that θ is distributed normally with mean 0. As detailed below, they learn its value gradually over time as information arrives.

The second component of the fully informed ideal policy, x_i for agent i , varies across the population. It reflects idiosyncratic aspects of each voter’s preferences. An important feature of our model is that agent i is also initially uncertain about x_i , and gradually learns its value over time as information arrives. The simplest interpretation of this uncertainty is that the agent’s preferences encompass idiosyncratic concerns that are informed by news. We prefer an alternative interpretation that offers a more compelling account of decisiveness: agent i ’s uncertainty concerning x_i reflects the “fuzziness” of i ’s preferences. Someone with sharp preferences (alternatively, a strong “moral compass”) understands in advance how she will weigh the various costs and benefits of the policy, and consequently knows the value of x_i with reasonable precision at the outset. In contrast, someone with fuzzy preferences (alternatively, a weak “moral compass”) has not yet arrived at definitive criteria for aggregating costs and benefits; she does not know in advance which weights she will ultimately use. Even when all consequences are known, such individuals will “waffle” as the arrival of extraneous information – for example, the subjective opinions of friends, advisors, constituents, and donors with vested interests – influences their overall evaluations. Returning again to the example of the financial crisis, the realization of x_i might depend on the particular way in which agent i weighs the various consequences of a bailout. It may be clear from the outset that a highly principled conservative will emphasize precedent and incentives over economic hardship, and that a highly principled liberal would do the opposite, but it is less clear where someone who lacks either a conservative or liberal “moral compass” will end up. In Appendix B, we provide formal microfoundations for this interpretation of our model.

We assume that the idiosyncratic preference of voter i , x_i , is distributed normally with mean μ_i (interpreted as political predisposition), and we index voters so that $i > j$ if $\mu_i > \mu_j$. We use f_μ to denote the commonly known population distribution of μ_i ; we assume it has full support over $[\underline{\mu}, \bar{\mu}]$ and is symmetric around 0. Agent i initially knows μ_i , but not x_i or θ . We use x_P to denote the idiosyncratic preference of a generic politician. At the outset, we assume for the sake of simplicity that all politicians share the same *ex ante* idiosyncratic preferences, μ_P . We introduce heterogeneity among politicians along this dimension in Section 6.

After taking office, politician P invests time τ gathering signals about θ and x_P , incurring perceived costs of $c_P\tau$, $c_P > 0$, to the politician and $c_i\tau$, $c_i > 0$, to voter i . We allow for the possibility that agents disagree about the costs of delay. The values of c_i are drawn independently from a distribution f_c ; we assume it has full support over the interval $[\underline{c}, \bar{c}] \subset \mathbb{R}_{++}$, is symmetric around the median, and is independent of μ_i .¹⁴ The realized cost for any particular agent is private information. The most natural interpretation of c_i is that it represents an opportunity cost. For example, delaying a decision may prolong the suffering caused by a crisis. An alternative interpretation is that c_i reflects the importance of choosing the right policy. Formally, a high value of c_i is equivalent to having relatively flat preferences over the policy space, a characteristic that reduces the perceived benefits of delay.¹⁵ Thus, we interpret a high value of c_i as reflecting high opportunity costs, low stakes (in the sense that the policy choice matters less), or both.

We assume that the variances of agent i 's initial forecast errors, $E_0[\theta^2]$ and $E_0[x_i - \mu_i]^2$, are α/ϕ and β_i/ϕ , respectively. As mentioned in the introduction, we allow for the possibility that politicians differ with respect to the parameter β_P .¹⁶ We model a politician with a strong “moral compass” as having a low value of β_P : because she knows in advance how she will ultimately weigh the various costs and benefits of the policy, she (and others) can gauge the idiosyncratic component of her preferences, x_i , with precision even at the outset. Such individuals display consistent political attitudes, in the sense that their preferred policies depart from the best current estimate of θ , either to the “right” or to the “left,” by predictable, stable margins. Conversely, we model a politician with a weak “moral compass” as having a high value of β_P : because she does not know in advance how she will ultimately weigh the various costs and benefits of the policy, she (and others) cannot gauge the idiosyncratic component of her preferences, x_i , with precision before knowing which extraneous influences she will encounter. Such individuals will display inconsistent political attitudes, in the sense that they will “waffle,” preferring policies that depart from the best current estimate of θ by highly variable margins. Again, formal microfoundations for this interpretation of our model appear in Appendix B. Let f_β denote the distribution of β among the population of politicians. We assume f_β is independent of μ and c and has full support over $[\underline{\beta}, \bar{\beta}] \subset \mathbb{R}_{++}$, where $f_\beta(\beta) \geq \underline{f}_\beta > 0$ and $f_\beta(\beta) \leq \bar{f}_\beta < \infty$.¹⁷

¹⁴Our distributional assumptions for μ and c are inessential but simplify some of the analytics.

¹⁵To appreciate the equivalence, simply renormalize the agent's utility function (which appears at the end of this section) by the multiplicative factor $1/c_i$.

¹⁶Whether or not we allow for heterogeneity of β_i for other agents is inconsequential.

¹⁷In principle, one could also introduce heterogeneity with respect to α . A lower value of α would indicate a greater technical aptitude for understanding complex policy questions. In other words, a person with a

Information gathering (which may include introspection about the decision maker’s own preferences) and deliberation improve the accuracy with which the politician P assesses θ and x_P . After investing time τ , the posterior distributions P holds about $E_\tau[\theta]$ and $E_\tau[x_P]$ are random variables that depend on the information P has gathered. We assume that at any point in time τ , P can collapse all information received about θ and x_P , respectively, into two normally distributed sufficient statistics, s^θ and s_P^x , the first with mean θ and variance α/τ , the second with mean x_P and variance β_P/τ .¹⁸ In that case, P ’s forecasts as of time τ are

$$E_\tau[\theta] = \frac{\tau}{\tau + \phi} s^\theta$$

$$E_\tau[x_P] = \frac{\phi}{\tau + \phi} \mu_P + \frac{\tau}{\tau + \phi} s_P^x.$$

It is straightforward to verify that the time τ forecast error for θ (that is, $\theta - E_\tau[\theta]$) is distributed normally with mean zero and variance $\frac{\alpha}{\phi + \tau}$, while the time τ forecast error for x_P (that is, $x_P - E_\tau[x_P]$) is distributed normally with mean zero and variance $\frac{\beta_P}{\phi + \tau}$. Thus, the variances of the time τ forecast errors asymptote to 0 as τ rises. As these formulas suggest, one can think of ϕ as parameterizing the amount of information contained in signals observed prior to $\tau = 0$.

3.2. Timing. In principle, politicians may signal decisiveness in two distinct types of settings. First, a politician who aspires to higher office may have opportunities to act decisively while holding lower office. Second, a first-term politician who aspires to win a second term may have opportunities to act decisively during her first term. In either case, opportunities

lower value of α would learn more (i.e., gather more information) about θ within any fixed information-gathering period. As we show (see equation (3), below), absent signaling, low- α candidates stop gathering information sooner, and it is straightforward to demonstrate that their information is more precise when they stop. (Specifically, the welfare loss due to imprecise knowledge is $(c_P(\alpha + \beta_P))^{1/2}$, which is increasing in α .) Therefore, if candidates differed with respect to α , then voters would unanimously prefer those with lower α and politicians would accelerate their decisions to signal technical competence. (A formal demonstration closely resembles our analysis of signaling with heterogeneity in β_P , which appears in Appendix C). We have drawn a distinction, however, between technical competence and decisiveness. For the reasons discussed in the introduction, we understand the statement “person A is more decisive than person B” to mean that A makes a decision more quickly than B when A and B have equal access to information. Politicians with low values of α make decisions more rapidly with *more* information. In contrast, politicians with high values of c_P , a focus of our analysis, make decisions quickly because they see delay as highly costly; hence, they make decisions more rapidly with *less* information.

¹⁸One can view these formulas as pertaining to the limiting case (as n goes to infinity) of a sequence of environments in which P starts out with ϕn iid normal signals and receives n additional signals per unit time, each with variance proportional to n . Focusing on the limiting case is analytically advantageous because the formulas become continuous in τ .

for decisiveness arise most frequently for executive officials such as first-term Presidents, governors and mayors.¹⁹ As a formal matter, we analyze settings of the first type, wherein officials signal decisiveness while holding lower offices (e.g., governors) in the expectation of running against each other for higher office (e.g., the Presidency). However, as we explain below, one can reinterpret our model as representing other settings, including those of the second type.

Our basic model consists of the following three stages.

In stage 1, all politicians hold (different) offices. Their preferences are as described in the preceding subsection. Each politician chooses the amount of time to deliberate and chooses a policy. In this stage, the strategic incentive to signal a type that voters prefer influences the politician's decisions. We assume that the politician takes the stage 1 action before having any information about the potential stage 2 election opponent.

In stage 2, two politicians compete in an election. Voters cannot directly observe the sharpness of a candidate's world view (β_P) or perceived costs of delay (c_P). Instead, the voters make inferences about those characteristics based on the candidate's decisions in stage 1. Voters have no other information about the candidates. We do not model the decision to seek higher office, and instead take the distributions of β_i and c_i among candidates (f_β and f_c) as given. In the equilibrium of our baseline model, the voters elect the more decisive of the competing politicians with certainty.

In stage 3, the politician who wins the stage 2 election assumes office, deliberates, and chooses a policy. Because there is no subsequent interaction, stage 3 decisions concerning p and τ are non-strategic. The payoff to the losing politician from stage 2 is the same as for a voter with the same characteristics.

To calculate a politician's total payoff, we sum across stages and attach weights of λ and unity to the stage 1 and stage 3 payoffs respectively.²⁰ A larger value of λ could reflect a belief that the stage 1 decisions are more important relative to stage 3 decisions, greater time discounting, or a longer time spent in lower office discounted at a fixed rate.

While we have modeled settings in which officials signal decisiveness while holding lower office in the expectation of running against each other for higher office, one can easily reinterpret our model to accommodate other settings. In our model, an officeholder anticipates running against an opponent whose decisiveness characteristics will be drawn from

¹⁹A legislator may be able to act decisively by quickly sponsoring new bills or publicly demanding swift and specific action in response to emerging problems. However, what qualifies as a decisive act is more nebulous since legislative action tends to involve the coordination of large groups of legislators and an often slow process of writing bills.

²⁰Recall that there is no immediate payoff in stage 2.

some known distribution and revealed to the electorate (through a separating equilibrium). However, the only role played by the opponent's decisiveness characteristics is that they determine the opponent's attractiveness to voters. Nothing of substance changes in our analysis if we replace each potential opponent with an alternative opponent who is equally attractive to the electorate in light of other characteristics, such as charisma. Thus, for example, in settings where the politician's objective is to signal only β (due to population homogeneity in c), we can reinterpret the value of β for a given opponent as the threshold value of β for the officeholder that would just allow the officeholder to defeat that opponent in light of the opponent's other characteristics. It follows that we can reinterpret our model as depicting a setting in which a particular official signals decisiveness while holding lower office (e.g., a governor) with the expectation of running for higher office (e.g., the Presidency) against someone with no opportunities to signal decisiveness (e.g., a member of the House). Similarly, we can reinterpret our model as depicting a setting in which an incumbent signals decisiveness (e.g., a first-term President) with the expectation of running for reelection against someone with no opportunity to signal decisiveness. The problem facing the signaler is essentially identical in all of these cases.²¹

We solve the model through backwards induction. Section 4 focuses on the politicians' behavior in stage 3 if they win the election and on assessing the endogenous valence component of voters' preference for decisive politicians. Section 5 studies how the rewards received by the elected politician in stage 3 define the incentives in stage 1 to signal a type that voters prefer and how the signaling behavior impacts the voters and the politicians. We introduce differences in politicians' policy preferences in Section 6, and demonstrate that more decisive politicians have greater ability to impose their agendas on the electorate.

4. VOTER PREFERENCE FOR DECISIVENESS

We begin by investigating voters' preferences over politicians' characteristics, assuming that politicians behave non-strategically when deliberating (as they will in stage 3). As a first step, we must determine how different types of politician would behave if elected. Given the quadratic loss function assumed in the previous section, P 's optimal policy choice after a deliberation period of length τ is

$$p^* = E_\tau[\theta] + E_\tau[x_P].$$

²¹We acknowledge that our analysis does not encompass situations in which a politician knows her future electoral opponent's attractiveness to voters at the point in time when she attempts to signal decisiveness. However, because voters' perceptions of candidates evolve over time in unpredictable ways, we regard situations where the signaler perfectly knows her future opponent's attractiveness as unimportant in practice.

Within each stage, we measure time from the start of that stage. As of time 0 in stage 3, the politician's expected utility is

$$\begin{aligned}
 W_P(\tau) &= E_0 [U_P(p, \tau)] \\
 &= -E_0 [(\theta - E_\tau[\theta]) + (x_P - E_\tau[x_P])]^2 - c_P\tau \\
 (1) \quad &= -\frac{\alpha + \beta_P}{\tau + \phi} - c_P\tau.
 \end{aligned}$$

The first term captures the costs of making a decision before learning the exact values of θ and x_P , and the second reflects the welfare loss from delay. Thus, the optimal length of the deliberation period is the solution to the following maximization problem

$$(2) \quad \max_{\tau \geq 0} -\frac{\alpha + \beta_P}{\tau + \phi} - c_P\tau.$$

From the first-order condition, we infer that

$$(3) \quad \tau^*(\beta_P, c_P) = \sqrt{\frac{\alpha + \beta_P}{c_P}} - \phi.$$

To ensure $\tau^* > 0$, we assume

$$\bar{c} < \frac{\alpha + \beta_P}{\phi^2}.$$

Voters' preferences over the politician's speed of deliberation are more complex. We can write the expected utility of voter i as of time 0 with politician P in office and a deliberation period of length τ as follows

$$\begin{aligned}
 (4) \quad W_i^P(\tau) &= E_0 [U_i(p, \tau) \mid \mu_P, \beta_P, c_P] \\
 &= -E_0 [(\theta - E_\tau[\theta]) + (x_i - E_\tau[x_P])]^2 - c_i\tau \\
 &= -\frac{\alpha}{\tau + \phi} - E_0 (x_i - E_\tau[x_P])^2 - c_i\tau.
 \end{aligned}$$

The first term represents the costs of making a decision before P learns the exact value of θ . The second terms captures the welfare loss due to the disparity between the voter's ideal and P 's expected idiosyncratic preferences as of time τ . The final term is the cost of delay.

To simplify equation (4), note that we can rewrite $E_\tau[x_P]$ as

$$E_\tau[x_P] = \frac{\phi}{\tau + \phi}\mu_P + \frac{\tau}{\tau + \phi}x_P + \frac{\tau}{\tau + \phi}\varepsilon_P^x,$$

where, by our assumptions, $\varepsilon_P^x \equiv s_P^x - x_P$ is normally distributed with mean 0 and variance β_P/τ . After some algebra one finds that

$$\begin{aligned} E_0 (x_i - E_\tau [x_P])^2 &= E_0 \left(x_i - \frac{\phi}{\tau + \phi} \mu_P - \frac{\tau}{\tau + \phi} x_P - \frac{\tau}{\tau + \phi} \varepsilon_P^x \right)^2 \\ &= \frac{\beta_i}{\phi} + \frac{\beta_P}{\phi} \frac{\tau}{\tau + \phi} + (\mu_i - \mu_P)^2. \end{aligned}$$

Putting all of these parts together yields:²²

$$(5) \quad W_i^P(\tau) = \underbrace{-(\mu_i - \mu_P)^2}_{\text{Policy Preference}} - \underbrace{\frac{\alpha}{\tau + \phi} - \frac{\beta_P}{\phi} + \frac{\beta_P}{\tau + \phi} - \frac{\beta_i}{\phi}}_{\text{Risk Preference}} - \underbrace{c_i \tau}_{\text{Cost of Delay}}.$$

The preceding expression decomposes the voter's preferences into a policy preference component, a risk preference component, and a component that reflects the cost of delay. The policy preference term captures the difference between the voter's and politician's preferred policy choices as of $\tau = 0$. The *risk preference* term reflects the implications of uncertainty concerning the voter's and politician's ideal policies. The *cost of delay* term captures the opportunity costs the politician imposes on the members of the public when deciding slowly.

We can further decompose the risk preference term into three components, as follows

$$(6) \quad \underbrace{-\frac{\alpha}{\tau + \phi}}_{\text{Risk from Common Concerns}} \quad \underbrace{-\frac{\beta_P}{\phi} + \frac{\beta_P}{\tau + \phi}}_{\text{Risk from Idiosyncratic Preference of the Politician}} \quad \underbrace{-\frac{\beta_i}{\phi}}_{\text{Risk from Idiosyncratic Preference of the Voter}}.$$

The first term captures risk from residual uncertainty about the common component of policy preferences. The second term reflects risk from the politician's efforts to fine-tune the policy based on how the fuzziness of his world view is resolved (i.e., the realization of x_P). When $\tau = 0$, this term is 0 because the politician has learned nothing about x_P beyond its a priori mean μ_P , and hence the voter bears no risk. As τ increases, the politician's views become sharper (i.e., the precision of his beliefs about x_P increases), and hence the voter bears more risk. As $\tau \rightarrow \infty$, this term approaches $-\frac{\beta_P}{\phi}$, which is the risk penalty imposed on the voters if the politician fully resolves the fuzziness of his world views. The final term captures the fuzziness of the voters' views about how to weigh the different aspects of the decision. Unlike the second term, it does not depend on τ : though voters may resolve the fuzziness of their views as time passes, that information does not

²²We will occasionally use the alternative form $W_i^P(\tau) = -(\mu_i - \mu_P)^2 - \frac{\alpha_P - \beta_P}{\tau + \phi} - \frac{\beta_i}{\phi} - \frac{\beta_P}{\phi} - c_i \tau$.

inform the politician's choice, and hence the welfare cost is the same regardless of when the politician chooses.

Our first proposition shows that a voter's favorite candidate perceives greater costs of delay than the voter ($c_P > c_i$) and has maximally sharp views (the smallest possible value of β_P). For any fixed τ , a preference for smaller values of β_P follows immediately from Equation 5, but the endogeneity of τ renders preferences over β_P more complex, so there is something more to prove. Preferences over the politician's costs of delay are more subtle.

Proposition 1. *From the perspective of voter i , the ideal politician has minimal uncertainty about his own preferences ($\beta_P = \underline{\beta}$) and perceives the cost of delay per unit time to be*

$$c_P = \begin{cases} c_i \frac{\alpha + \underline{\beta}}{\alpha - \underline{\beta}} & \text{if } \alpha > \underline{\beta} \text{ and the result is less than } \bar{c} \\ \bar{c} & \text{otherwise} \end{cases}.$$

The intuition for this result is straightforward. Voter i does not benefit from any information P learns that helps resolve the fuzziness of his views, x_P . On the contrary, such information adds risk from i 's perspective. Consequently, i would ideally like to elect a politician P^* with $\beta_{P^*} = 0$ and $c_{P^*} = c_i$, who would spend time $\tau^*(0, c_i)$ making a decision.²³ Of course, with $\underline{\beta} > 0$, i 's ideal politician doesn't exist. The best available choice for i is to select a politician P' with $\beta_{P'} = \underline{\beta}$. To reduce the risk associated with P' learning about $x_{P'}$, i would like P' to make the decision even more rapidly than P^* (so $\tau^*(\underline{\beta}, c_{P'}) < \tau^*(0, c_i)$). With $c_{P'} \leq c_i$, P' would actually take *more* time than P^* (because then we would have $\tau^*(\underline{\beta}, c_{P'}) > \tau^*(0, c_i)$ by equation (3)). Consequently, it must be the case that $c_{P'} > c_i$, as the proposition implies.

Notice that if voter i were to become a politician and make the decision, i would take more time than P^* in order to learn about x_i (i.e., $\tau^*(\beta_i, c_i) > \tau^*(0, c_i)$). Hence i would also take more time than i 's most preferred politician P' (because $\tau^*(\underline{\beta}, c_{P'}) < \tau^*(0, c_i)$). Since all voters favor politicians who make decisions faster than themselves, a majority of the population would like to elect faster-than-average leaders who "know their own minds" ($\beta_P = \underline{\beta}$) and perceive greater-than-average costs of delay. Thus we begin to arrive at an explanation for the electoral success of decisive politicians.

As a general matter, existence of a Condorcet winner can be a problematic issue, particularly in settings with more than one dimension of heterogeneity as we have in this model (even given our provisional assumption that all politicians share the same ex ante

²³Recall that if $\beta_{P^*} = 0$, then the politician knows his own mind perfectly. Since the politician's views are not at all fuzzy, the voters suffer no risk from the politician tuning his policy to what he learns about x_P as there is nothing to learn.

idiosyncratic preferences, μ_P). Fortunately, the model has three special properties that simplify matters. First, voter i 's utility depends on c_P only through the decision time, τ . We can therefore think of each politician as offering an alternative in (τ, β_P) -space rather than in (c_P, β_P) -space. Second, as shown in the proof of Proposition 1, preferences are single-peaked in τ for any fixed β_P . Third, as is clear from equation (5), all voters prefer politicians with lower values of β_P , meaning we can limit ourselves to considering potential Condorcet winners of the form $(\tau, \underline{\beta})$. Existence of a Condorcet-winning candidate then follows immediately: pairing the median preferred value of τ with $\underline{\beta}$, we identify a candidate who majority defeats all other $(\tau, \underline{\beta})$ combinations. From Proposition 1, it then follows that the value of c_P for the Condorcet-winning candidate is greater than the population median. Hence the winning candidate will make decisions faster than the typical member of the community.

Formally, we define decisiveness in terms of the speed with which the politician makes decisions relative to τ_{med} , the population median of $\tau^*(\beta_i, c_i)$:

Definition 1. *A politician of type (β_P, c_P) is **decisive** if $\tau^*(\beta_P, c_P) < \tau_{med}$.*

The preceding discussion then yields the following corollary of Proposition 1, which formalizes our first main point by showing that majoritarian institutions tend to select decisive politicians.

Corollary 1. *There exists a Condorcet winner with $\beta_P = \underline{\beta}$ and*

$$c_P = \begin{cases} c_{med} \frac{\alpha + \beta_P}{\alpha - \beta_P} & \text{if this quantity is less than } \bar{c} \\ \bar{c} & \text{otherwise} \end{cases},$$

where c_{med} is the median of the distribution of c . Moreover, this candidate is decisive.

Instead of evaluating a politician's decision speed relative to that of the typical citizen, one could compare it to the speed with which any given voter i would like the politician to make a decision, $\tau^v(\beta_P; c_i)$. In the proof of Proposition 1, we showed that

$$(7) \quad \tau^v(\beta_P; c_i) = \sqrt{\frac{\alpha - \beta_P}{c_i}} - \phi.$$

Notice that τ^v depends on β_P : from the voter's perspective, P 's deliberation introduces more risk when β_P is larger.

Definition 2. *Voter i considers politician P **hesitant** if P 's decision time is greater than $\tau^v(\beta_P; c_i)$ and **hasty** if it is less than $\tau^v(\beta_P; c_i)$.*

Given that $\beta_P = \underline{\beta}$ is a property of the politicians and hence fixed across voters, a voter's preference over politicians is determined by the difference between the voter's and politician's aversion to delay, which yields voter preferences over c_P that are single-peaked (holding $\beta_P = \underline{\beta}$ fixed). If we have $c_{\text{med}} \frac{\alpha + \beta_P}{\alpha - \beta_P} \leq \bar{c}$ (i.e., we are not in the corner case of Corollary 1), then we find that exactly half the population considers the Condorcet-winning candidate hesitant, while the other half considers that candidate hasty. In other words, a majority of voters would oppose either an increase or a decrease in the candidate's decision speed.

5. SIGNALING DECISIVENESS

In this section we explore the idea that politicians who expect to seek higher office (stage 2 of our model) may attempt to cultivate reputations for decisiveness by accelerating decisions made in lower offices (stage 1 of our model). The implications of our analysis differ according to whether politicians wish to signal an aversion to delay (high c) or the sharpness of their world views (low β). In the main text we focus on signaling an aversion to delay as this proves to be the more interesting case – for example, it implies that the median voter's favorite politician type earns the lowest utility in equilibrium. We analyze signaling the sharpness of world views in the online appendix. Throughout we focus on fully separating equilibria in which the voters infer (correctly in equilibrium) the types of the politicians from the politician's choices in stage 1.

For most of this section we will assume the median voter's preferences are monotonic in c . Either $\alpha < \beta$ or $c_{\text{med}} \frac{\alpha + \beta}{\alpha - \beta} > \bar{c}$ suffices to guarantee this property; both conditions ensure that any increase in c up to \bar{c} benefits the median voter by reducing the politician's desire to resolve the fuzziness of his world view and, as a result, impose risk on the voters. At the end of this section, we comment briefly on the case in which the median voter's ideal lies on the interior of $[\underline{c}, \bar{c}]$. We will also continue to assume for simplicity that candidates share the same *ex ante* policy preferences, μ_P . At the cost of carrying around some extra terms, one can easily extend this analysis to subsume known differences in *ex ante* preferences, reflecting for example the differing ideologies of two political parties. One can also incorporate the types of ideological compromises we introduce in the next section without qualitatively altering our conclusions (as shown in an earlier version of this paper).

5.1. Separating equilibrium. Consider the stage 1 decision problem facing a politician with a perceived cost of delay c who expects to play the role of a candidate in the stage 2 election. From the preceding discussion, the candidate knows he will win the election if voters believe he has a greater aversion to delay than his opponent. Because we are

studying fully separating equilibria, the stage 1 outcome will fully reveal both candidates' types. Therefore, when choosing the image he wishes to project in stage 1, he knows he will win only if the voters believe his cost of delay is greater than his opponent's actual cost of delay.

In effect, a fully separating equilibrium presents each candidate with a menu of feasible action-perception pairs, where the action is stage 1 decision time and the perception pertains to the politician's aversion to delay. For our purposes it is convenient to represent this schedule as a function $\tau^S : [\underline{c}, \bar{c}] \rightarrow \mathbb{R}_+$. One can think of each candidate as choosing the desired perception \hat{c} , and then making a stage 1 decision after deliberating for the amount of time required to create that perception, $\tau^S(\hat{c})$. The electoral victor will be the candidate whom voters believe is most averse to delay (that is, for whom \hat{c} is greatest). As mentioned before, with full separation, each candidate knows he will win if his chosen \hat{c} is greater than his opponent's true cost-of-delay parameter. Accordingly, we can write his stage 3 payoff as follows

$$(8) \quad \begin{aligned} \Pi(c, \hat{c}) = & - \int_{\underline{c}}^{\hat{c}} \left[\frac{\alpha + \beta}{\tau^*(c) + \phi} + c\tau^*(c) \right] f_c(s) ds \\ & - \int_{\hat{c}}^{\bar{c}} \left[\frac{\alpha - \beta}{\tau^*(s) + \phi} + 2\frac{\beta}{\phi} + c\tau^*(s) \right] f_c(s) ds. \end{aligned}$$

The first integral reflects the politician's payoff when he wins the election,²⁴ and the second is the politician's payoff when he loses the election to an opponent with a higher perceived aversion to delay.

How do stage 3 payoffs vary with the inference, \hat{c} ? Taking the derivative, we have:

$$(9) \quad \left. \frac{\partial \Pi(c, \hat{c})}{\partial \hat{c}} \right|_{\hat{c}=c} = \frac{2\beta}{\phi} \frac{\tau^*(c)}{(\tau^*(c) + \phi)} f_c(c).$$

The right hand side of equation 9 represents the benefit of winning an election against an opponent with the same aversion to delay. We interpret this term as an endogenous rent from holding office. These rents include (1) the benefit of fine-tuning the policy based on the resolution of a fuzzy world view and (2) the avoidance of the risk associated with another candidate doing the same.

²⁴Since we are integrating a constant function, we could simply write this as the probability of winning times the payoff in the event he wins. We write the integral explicitly for easy comparison to the payoff in the event the politician loses and to facilitate a comparison with the extensions to electoral competition with policy commitments in Section 6.

Including stage 1, a politician's total payoff is

$$V(c, \hat{c}) = \Pi(c, \hat{c}) - \lambda \left[\frac{\alpha + \beta}{\tau^S(\hat{c}) + \phi} + c\tau^S(\hat{c}) \right].$$

To determine the candidate's optimal choice of \hat{c} given the signaling schedule τ^S , we take the derivative of V with respect to \hat{c} and set it equal to zero: $\frac{\partial}{\partial \hat{c}} V(c, \hat{c}) = 0$. In a fully separating equilibrium the solution is $\hat{c} = c$, which implies $\frac{\partial}{\partial \hat{c}} V(c, c) = 0$, or equivalently

$$(10) \quad \left. \frac{\partial \Pi(c, \hat{c})}{\partial \hat{c}} \right|_{\hat{c}=c} = \lambda c \left[1 - \left(\frac{\tau^*(c) + \phi}{\tau^S(c) + \phi} \right)^2 \right] \frac{\partial \tau^S}{\partial \hat{c}}.$$

Equation 10 follows from rearranging the first order condition, $\partial V / \partial \hat{c} = 0$, and using the fact that $\tau^*(c) + \phi = \sqrt{(\alpha + \beta)/c}$.

Observe that equation 10 is a nonlinear first-order differential equation. As usual, the equilibrium leaves the choice of the "worst" type undistorted, so we also have a boundary condition, $\tau^S(\underline{c}) = \tau^*(\underline{c})$. A somewhat unconventional feature of this differential equation is that the coefficient of $\frac{\partial \tau^S}{\partial c}$ is zero at the initial condition, which renders $\left. \frac{\partial \tau^S}{\partial c} \right|_{c=\underline{c}}$ undefined. We can finesse this difficulty by reversing the mathematical roles of τ^S and c , treating equation 10 as a differential equation for the function $c^S(\tau)$, with initial condition $c^S(\tau) = \underline{c}$ for $\tau = \tau^*(\underline{c})$. In that case, we have $\left. \frac{\partial c^S}{\partial \tau} \right|_{\tau=\tau^*(\underline{c})} = 0$. Our next result describes some properties of the solution.

Proposition 2. $c^S(\tau)$ is strictly decreasing in τ with $\tau < \tau^*(c^S(\tau))$ for $\tau < \tau^*(\underline{c})$ and strictly increasing in τ with $\tau > \tau^*(c^S(\tau))$ for $\tau > \tau^*(\underline{c})$.

This proposition offers us two candidates for the separating function τ^S : we can invert the solution c^S restricting attention either (1) to $\tau < \tau^*(\underline{c})$, which yields a downward-sloping function $\tau^1(c)$ satisfying $\tau^1(c) < \tau^*(c)$, implying that greater delay aversion translates into greater speed and signaling causes leaders to accelerate decisions or (2) to $\tau > \tau^*(\underline{c})$, which yields an upward-sloping function $\tau^2(c)$ satisfying $\tau^1(c) > \tau^*(c)$, implying the opposite. Given our assumptions, there is no guarantee that either solution is globally incentive-compatible. Our next result tells us that only τ^1 can serve as the separating function τ^S . We show that a sufficiently large value of λ guarantees global incentive compatibility of τ^1 .²⁵ In contrast, τ^2 is never globally incentive compatible. Accordingly, we will henceforth associate τ^S with τ^1 .

²⁵One might have hoped that, as in many signaling models, a supermodularity (or single-crossing) property would hold for $V(c, \hat{c})$, as standard techniques for proving incentive compatibility would then apply. Unfortunately, that is not the case.

Proposition 3. *There exists λ^* such that for $\lambda > \lambda^*$, τ^1 is globally incentive-compatible. τ^2 is never globally incentive-compatible.*

Together, Propositions 2 and 3 imply that signaling an aversion to delay accelerates decision making ($\tau^S(c) < \tau^*(c)$) – politicians act more decisive than they actually are. In addition, we learn that τ^S must be strictly decreasing in c , which raises the possibility that the non-negativity constraint ($\tau > 0$) may bind. We will eliminate this possibility by assuming, where necessary, that ϕ is sufficiently small.

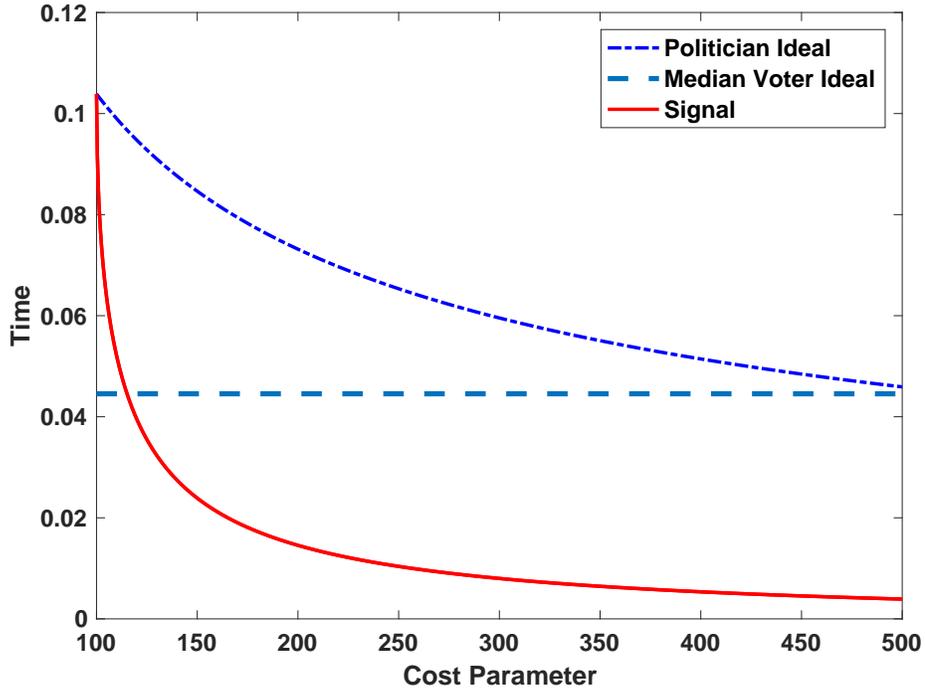


FIGURE 1. Equilibrium Versus Voter and Politician Optima With Signaling of Cost

Figure 1, which is based on a numerical simulation, depicts a typical equilibrium.²⁶ The horizontal axis describes the politician’s type, and the vertical axis describes the time taken to make a decision. The figure shows three functions relating the decision time, τ , to the politician’s delay aversion, c : the ideal from the politician’s perspective (τ^*), the median

²⁶The figure was generated using $\phi = 0.001$, $\alpha = 1$, $\beta = 0.1$, $\underline{c} = 100$, and $\bar{c} = 500$. c is distributed as per a truncated normal distribution with a mean of 500 and a standard deviation of 4,000. The example was checked for global incentive compatibility numerically.

voter’s ideal decision time (τ^v), and the signaling equilibrium (τ^S). Because α and β are common across all agents, each voter has a single decision time they would prefer the politician to choose regardless of the politician’s cost (see equation 7).²⁷

5.2. Welfare. We have seen that voters regard non-strategic politicians as hesitant and that signaling increases decision speed. One might therefore think that signaling would benefit voters. In fact, that is not necessarily the case. Signaling is a rather poor solution for politicians’ tendency to delay because the signaling incentive is smallest where the need for a corrective influence is greatest and greatest where that need is smallest. That pattern is evident from Figure 1. Signaling has no effect on politicians with the lowest values of c , for whom the gap between the voter and politician ideal is greatest. Moreover, its cumulative effect on politicians with the highest values of c , for whom the gap between τ^* and τ^v is smallest, can be enormous, causing them to spend little or no time pondering the common good even when they should. In the figure, the signaling curve is so steeply sloped that it crosses the voter ideal curve. Politicians to the left of the crossing remain hesitant from the voter’s perspective, but those to the right become hasty. Thus the overall impact on voter welfare can be positive or negative, depending on the size of the signaling distortion and the distribution of politician types – specifically, whether it is skewed toward those with relatively high aversions to delay who overcorrect, or those with relatively low aversions to delay who undercorrect. That said, a “little bit” of signaling (sufficiently large λ) unambiguously improves welfare: when the signaling curve is sufficiently close to the politician ideal curve, it lies between the politician ideal and voter ideal curves.

These results have potentially interesting implications for the role of incumbency in electoral politics. Incumbents with a high value of c that have fully revealed their types, or who have reached their final terms under term-limited regimes, have sharply attenuated signaling motives (i.e., they are not hasty). Thus, in cases such as the one shown in Figure 1, the theory can produce an endogenous preference for incumbents.

5.3. Extensions.

5.3.1. Endogenous candidate selection. As we saw in the previous section, the welfare effects of signaling can depend critically on the distribution of candidate types. While we have not endogenized that distribution, our model allows us to provide some useful insights. Our next result establishes that the voters’ least favored politicians receive higher

²⁷Recall that the politician’s cost is irrelevant for determining the length of time a voter would like the politician to take to make a decision.

expected payoffs in the separating equilibrium. Technically, this stems from the fact that voters prefer high cost politicians.

Proposition 4. $V(c, c)$ is decreasing in c .

Now imagine endogenizing the distribution of candidate types by appending the familiar citizen-candidate apparatus to our model (Besley and Coate [6]). Proposition 4 suggests that the distribution of politicians would be skewed toward types with low values of c (i.e., those who are least aligned with the voters' preferences). As a result, the effect of signaling would tend to be relatively small but beneficial, and voters would be more likely to complain that politicians are too hesitant.

5.3.2. *The effect of transparency.* Politicians generally make many decisions while in office. We can classify a political institution as more or less transparent according to whether voters obtain information about the deliberations associated with a large or small fraction of these decisions. In this section, we examine the effect of changes in transparency on political outcomes. Intuitively, transparency leads politicians to “spread” their signals across many decisions. As a result, each choice is subject to less distortion, and deliberation times are closer to the politicians' ideals. However, because politicians and voters have different ideals, transparency can hurt voters' interests.

Formally, suppose the politician makes N stage 1 decisions rather than one. Each has the structure described in Section 3, and the N stochastic realizations are independent. In $M \leq N$ instances, the electorate observes the deliberation process. For the purpose of studying transparency, we are interested in the effects of changing M while holding N fixed.

We denote the stage 1 actions as $\tau^S = (\tau_1^S, \dots, \tau_N^S)$. There are, of course, many ways to signal a single characteristic through multiple actions, and we consider fully separating equilibria with continuous action mappings $\tau_m^S : [c, \bar{c}] \rightarrow \mathbb{R}_+$ for $m = 1, \dots, N$, where voters observe τ_m for $m \leq M$. Note that in any such equilibrium, unobserved choices are undistorted: a politician with a perceived cost of delay c sets $\tau_m = \tau^*(c)$ for $m = M + 1, \dots, N$.

The simplest equilibrium within this broad class treats all observable tasks symmetrically: $\tau_m^S = \underline{\tau}^0$ for $m = 1, \dots, M$. With this restriction, the fully separating action function is the solution to the following differential equation

$$(11) \quad \left. \frac{\partial \Pi(c, \hat{c})}{\partial \hat{c}} \right|_{\hat{c}=c} = M\lambda c \left[1 - \left(\frac{\tau^*(c) + \phi}{\tau^S(c) + \phi} \right)^2 \right] \frac{\partial \underline{\tau}^0}{\partial c},$$

with boundary condition $\underline{\tau}^0(\underline{c}) = \tau^*(\underline{c})$. This is, of course, a slightly modified version of Equation 10.

How will politicians choose to signal their decisiveness? Will they do so through patterns of consistently decisive behavior, or will they occasionally display extreme decisiveness? We take the view that they will seek, and over time discover, the most efficient ways to signal.²⁸ Our next result establishes that this process drives them toward the symmetric equilibrium in which they are consistently decisive.

Proposition 5. *The symmetric equilibrium maximizes the payoff for every type of politician within the set of fully separating equilibria.*

As the proof demonstrates, Proposition 5 holds because symmetry magnifies the rate at which a greater aversion to delay reduces the welfare loss from making decisions too quickly. It is unrelated to concavity of the representative politician's utility function.

Our final result shows that decision times converge to the politicians' ideals when politicians make many decisions that are observable to voters.

Proposition 6. $\underline{\tau}^0(c)$ converges to $\tau^*(c)$ uniformly as $M \rightarrow \infty$.

This result suggests that the desire to cultivate a reputation for decisiveness will distort any given deliberation to a smaller degree when the politician occupies a position that provides many opportunities to make visible decisions. Compare, for example, the signaling opportunities available to governors and legislators. As an executive, a governor has many opportunities to demonstrate decisiveness, but as a member of a larger deliberative body, a legislator likely has few. According to our theory, the legislator may feel compelled to act with extreme haste when opportunities for independent action arise. For example, she might craft and sponsor a bill in response to some emergent issue without adequate consideration or vetting.²⁹ In contrast, the governor might be in a position to act less precipitously without compromising his reputation.

In our model, the effect of greater transparency on voter welfare is, as a general matter, ambiguous. In some cases it is beneficial because it reigns in hasty decision making. In others it is harmful because it slows down hesitant decisions. With sufficiently high levels of transparency, we have $\underline{\tau}^0(c) > \tau^v(c)$ and signaling necessarily becomes welfare-improving because it no longer causes politicians to overshoot the voter ideal. However, incremental

²⁸Plausible restrictions on out-of-equilibrium beliefs often point to the efficient separating equilibrium; see, for example, Cho and Kreps [8] or, for a problem with a more similar structure, Bernheim [3].

²⁹We acknowledge that legislators may act hastily for other reasons, for example to set the legislative agenda.

transparency beyond that point is harmful because it causes deliberation times to approach $\tau^*(c)$, which all voters regard as hesitant. From the perspective of institutional design, natural objectives therefore come into conflict: on the one hand, transparency promotes accountability; on the other hand, for the reasons we have discussed, it may exacerbate other agency problems between voters and office holders.

5.3.3. *Non-monotonic voter preferences.* The case in which the delay aversion of the median voter's ideal politician lies on the interior of $[\underline{c}, \bar{c}]$ is considerably more complex. Politicians with low c will wish to project greater delay aversion than they feel, and those with high c will do the opposite. One can show that these opposing inclinations rule out the existence of fully separating equilibria. Instead, as in Bernheim [3] and Bernheim and Severinov [4], they can generate equilibria with a single pool at a point in the interior of $[\underline{c}, \bar{c}]$, generally near the median voter's ideal point. Moreover, depending on the model's parameters, the fraction of politicians who join this pool may be arbitrarily close to unity. In such cases, voters would observe uniformly high quality decision making by politicians in lower office, only to observe that quality deteriorates systematically, but to a highly varying degree, when politicians reach higher office.

6. HETEROGENEOUS POLICY PREFERENCES AND IDEOLOGICAL COMPROMISES

In this section, we extend the model to encompass the possibility that politicians have ex ante heterogeneous policy preferences and they can commit themselves in advance to ideological compromises. Because perceived decisiveness confers an electoral advantage, it enhances a leader's ability to impose his or her own agenda on voters. Greater political polarization magnifies this benefit, and therefore amplifies the incentive to appear decisive by rushing through deliberations.

6.1. **The extended model.** We assume the stage 2 election involves two candidates whose ex ante policy ideals lie equally far from the population median, but in opposite directions: $\mu_1 = -\mu_2 \equiv \mu = E[x_{P_1}]$, where subscripts indicate the candidate. Voters observe these ideological orientations. As in the previous section, we will also assume that politicians' world views, β , are equally sharp. One can relax these assumptions at the cost of extra notational complexity.

It would be unrealistic to assume that politicians would commit to inflexible policies before learning anything about the state of the world. After all, it is not in the interest of the politician or the electorate to make uninformed policy choices. Instead, we assume

that politicians can make commitments that are contingent on the information they subsequently acquire. One can think of these contingent commitments as political alliances that are intended to offset or reinforce the candidate's ideological biases, once elected. For example, they could involve the choice of key advisors, such as hawkish foreign policy experts, and/or commitments to provide particular interest groups with "seats at the table." Whenever we refer to a policy commitment, we have in mind a commitment that is contingent on the information the politician acquires concerning his ideal policy, $E_\tau[\theta] + E_\tau[x_P]$.

To keep matters relatively simple, we employ a reduced-form representation of this process. Specifically, in stage 2, politician P commits to an ideological adjustment, δ , which shifts the policy chosen after a deliberation period of length τ from $p^* = E_\tau[\theta] + E_\tau[x_P]$ to $p^* = E_\tau[\theta] + E_\tau[x_P] + \delta$. Setting $\delta \neq 0$ is costly to the politician because it shifts the ultimate policy (if he is elected) away from his ideal point. However, because δ is observable, this strategy may increase his chance of electoral victory.

Due to technical issues bearing on the existence of best responses, we assume that, in the event of an electoral tie, the winner is the candidate who appears more attractive to the majority of voters based on decision speed rather than policy predisposition. If a majority favors neither candidate based on this characteristic, each wins with probability $1/2$. These ties are probability 0 events in equilibrium.

6.2. Electoral equilibrium. For the purpose of this section, we focus on stages 2 and 3 of the model, and assume voters believe the candidates' characteristics are (μ_1, \hat{c}_1) and (μ_2, \hat{c}_2) (recall that we are assuming away heterogeneity in β). Our objective is to determine the equilibrium values of the ideological adjustments, δ_1 and δ_2 , as well as the expected rewards to each politician. As we will see, the winning candidate is the one who appears more attractive to the majority of voters based on decisiveness. That advantage enables the candidate to prevail despite making a smaller ideological concession than the rival. Because the office holder can implement a policy closer to his ideal, winning benefits the winner above and beyond the endogenous rents from holding office that are present in our original model.

An ideological compromise of δ reduces the politician's expected stage 3 payoff conditional on a deliberation period of length τ by the fixed amount δ^2 . It follows that ideological compromises do not alter $\tau^*(\beta_P, \hat{c}_P)$, the length of the politician's optimal stage 3 deliberations. As a result, a voter's expected payoff from electing a politician with characteristics (μ_P, \hat{c}_P) who commits to a compromise of δ is precisely the same as from electing a politician with characteristics $(\mu_P + \delta, \hat{c}_P)$ who makes no commitment. With this adjustment, we can continue to apply the formulas for voter preferences derived in Section 4.

The incremental benefit voter i receives from electing candidate 1 rather than candidate 2 is

$$\begin{aligned}\Delta_i &= W_1^P - W_2^P \\ &= 2\mu_i(2\mu + \delta_1 - \delta_2) - (\tau_1 - \tau_2)c_i + \\ &\quad \left[(\delta_2 - \mu)^2 - (\delta_1 + \mu)^2 + \left(\frac{\alpha - \beta}{\tau_2 + \phi} - \frac{\alpha - \beta}{\tau_1 + \phi} \right) \right],\end{aligned}$$

where τ_j refers to the anticipated speed of candidate j if he serves in office in stage 3. We will use M to denote the voter whose characteristics $\mu_M = 0$ and c_M are each population medians (henceforth, the “median voter”). Consider the set of voters who agree with M about the net value of electing candidate 1: $\Delta_i = \Delta_M$. If $-\delta_1 = \delta_2 = \mu$ and $\tau_2 = \tau_1$, this set includes all voters. Otherwise, $\Delta_i = \Delta_M$ is a linear relation dividing the space of voters into two half spaces according to whether $\Delta_i \leq \Delta_M$. Under our assumptions (independence and symmetry of the distributions of μ_i and c_i), those half spaces have equal mass. It follows that a (weak) majority of voters always prefers the same candidate as the median voter. Thus we have:

Lemma 1. *If a voter with characteristics (μ_M, c_M) strictly prefers candidate i to candidate j , then so does a majority of voters.*

Lemma 1 establishes that the outcome of the election depends on the preferences of the median voter, M . A little algebra reveals that candidate 1 wins with certainty if

$$(12) \quad (\mu + \delta_1)^2 - (\mu - \delta_2)^2 < \Phi_M(\hat{c}_1, \hat{c}_2),$$

where

$$\Phi_M(\hat{c}_1, \hat{c}_2) = \left(\frac{\alpha - \beta}{\tau(\hat{c}_2) + \phi} + c_M \tau(\hat{c}_2) \right) - \left(\frac{\alpha - \beta}{\tau(\hat{c}_1) + \phi} + c_M \tau(\hat{c}_1) \right).$$

Φ_M captures the component of the median voter’s preference attributable to the candidates’ decisiveness. If $\Phi_M > 0$, candidate 1 can win the election with an anticipated policy choice that is further from the median voter’s ideal than candidate 2’s anticipated choice. The following lemma tells us that a better reputation for decisiveness (defined in terms of proximity to the median voter’s ideal) puts a candidate in a better position to win the election.

Lemma 2. *$\frac{\partial}{\partial \hat{c}_1} \Phi_M(\hat{c}_1, \hat{c}_2)$ has the same sign as $c_M \frac{\alpha + \beta}{\alpha - \beta} - \hat{c}_1$, and $\frac{\partial}{\partial \hat{c}_2} \Phi_M(\hat{c}_1, \hat{c}_2)$ has the same sign as $\hat{c}_2 - c_M \frac{\alpha + \beta}{\alpha - \beta}$.*

The preceding analysis implies that the likelihood of winning shifts discontinuously from 0 to 1 as candidate i varies δ_i through the point at which $(\mu + \delta_1)^2 - (\mu - \delta_2)^2 = \Phi_M$.

This property generates Downsian convergence with respect to the ideological adjustments δ_1 and δ_2 . In equilibrium, the candidate who is less attractive to the median voter in terms of decisiveness (that is, the one with the larger value of $\frac{\alpha-\beta}{\tau_i+\phi} + c_M\tau_i$) perfectly aligns with the median voter through an ideological adjustment, but a majority votes for the other candidate, whose ideological adjustment is just sufficient to make the median voter indifferent. Accordingly, we have:

Proposition 7. *The Nash equilibria depend on the model's parameters as follows:*

- (1) *If $\mu^2 \geq \Phi_M > 0$, the unique pure strategy Nash equilibrium involves $\delta_1 = \sqrt{\Phi_M} - \mu$, $\delta_2 = \mu$, and the election of candidate 1.*
- (2) *If $\Phi_M > \mu^2$, all pure strategy Nash equilibria involve $\delta_1 = 0$ and the election of candidate 1.*
- (3) *If $\mu^2 \geq -\Phi_M > 0$, the unique pure strategy Nash equilibrium involves $\delta_2 = \mu - \sqrt{-\Phi_M}$, $\delta_1 = -\mu$, and the election of candidate 2.*
- (4) *If $-\Phi_M > \mu^2$, all pure strategy Nash equilibria involve $\delta_2 = 0$ and the election of candidate 2.*
- (5) *If $\Phi_M = 0$, then $\delta_1 = \delta_2 = -\mu$ and each candidate is elected with probability $\frac{1}{2}$.*

Notice how δ_i varies with candidate i 's relative decisiveness. When i 's decisiveness advantage is sufficiently large, i wins the election without compromising (cases 2 and 4). As the advantage declines, i continues to win, but makes larger and larger concessions in the direction of the median voter (cases 1 and 3). Once the advantage shifts to the competitor, i loses despite compromising fully with the median voter. Thus, more (appropriately) decisive candidates successfully impose their own agendas on the electorate.

The ability of (appropriately) decisive candidates to resist compromise and thereby earn greater rents from office-holding provides additional incentives for candidates to seek reputations for decisiveness. In the next section, we investigate the manner in which signaling is affected by the parameter μ , which measures the degree of partisan disagreement.

6.3. Signaling and the degree of partisanship. Conventional wisdom holds that the U.S. electorate has become increasingly polarized.³⁰ The degree of partisanship also differs sharply across different types of elections. For instance, it is less pronounced when the main competition for an office occurs between members of the same party (as is the case for many seats in the U.S. House of Representatives) than when it involves candidates from opposing parties (as is the case for U.S. Presidential elections). In this section, we show

³⁰The available evidence points to sharp polarization among the political elite. However, the evidence on mass polarization is less clear; see Fiorina and Abrams [13].

that politicians signal decisiveness more aggressively in regimes with greater partisanship. The intuition is straightforward: greater partisanship increases the net gains from winning an election and therefore increases the incentive to cultivate a reputation for decisiveness.

As in Section 5, we assume that the median voter's preferences are monotonic in the politician's perceived costs of delay. For the reasons discussed in Section 5.3.3, this simplification keeps the analysis of signaling tractable.

With ex ante heterogeneous policy preferences, the expected utility of a type c politician who is viewed as type \hat{c} becomes

$$(13) \quad \Pi(c, \hat{c}) = - \int_{\underline{c}}^{\hat{c}} \left[\left(\mu - \sqrt{\Phi_M(\hat{c}, s)} \right)^2 + \frac{\alpha + \beta}{\tau^*(c) + \phi} + c\tau^*(c) \right] f_c(s) ds \\ - \int_{\hat{c}}^{\bar{c}} \left[\left(\mu + \sqrt{-\Phi_M(\hat{c}, s)} \right)^2 + \frac{\alpha - \beta}{\tau^*(s) + \phi} + 2\frac{\beta}{\phi} + c\tau^*(s) \right] f_c(s) ds.$$

Provided one redefines $\Pi(c, \hat{c})$ in this way, equation 10 still describes the fully-separating equilibrium, and Propositions 2, 3, and 4 continue to hold with relatively minor changes to the proofs.

Although μ has no effect on a politician's marginal costs of signaling higher delay aversion (stage 1 payoffs), it changes the marginal benefits as follows:

$$(14) \quad \frac{\partial^2 \Pi(c, \hat{c})}{\partial \mu \partial \hat{c}} = \int_{\underline{c}}^{\hat{c}} \left[\frac{1}{\sqrt{\Phi_M(\hat{c}, s)}} \frac{\partial \Phi_M(\hat{c}, s)}{\partial \hat{c}} \right] f_c(s) ds \\ + \int_{\hat{c}}^{\bar{c}} \left[\left(\frac{1}{\sqrt{-\Phi_M(\hat{c}, s)}} \frac{\partial \Phi_M(\hat{c}, s)}{\partial \hat{c}} \right)^2 \right] f_c(s) ds > 0.$$

The inequality in Equation 14 implies that the marginal benefit of an increase in perceived aversion to delay (that is, an increase in \hat{c}) is increasing in μ , the degree of partisanship. Intuitively, partisan differences increase the stakes associated with winning and losing, and hence make politicians more willing to incur costs that will help them reach office.

Proposition 8. *In settings where politicians signal aversion to delay, $\tau^S(c)$ is decreasing in μ .*

As a practical matter this result suggests, for example, that politicians will cultivate decisive images more aggressively while in lower office if they aspire to higher offices for which elections are more highly partisan (for example, U.S. President rather than a Representative from a district closely allied with one party).

7. CONCLUSION

We have presented a theory that rationalizes voters' preferences for decisive leaders who reach decisions expeditiously despite possessing no special skill at information gathering or processing. In the setting we have studied, policy preferences have both a common component and an idiosyncratic component. Uncertainty concerning the latter component reflects the initial fuzziness of the politician's views, for example concerning the relative weighting of various costs and benefits. Office holders can take time to learn about both, but at a cost. For a politician with sharp preferences (a strong "moral compass"), the initial uncertainty is low. We have demonstrated that agency problems between voters and politicians create natural preferences among voters for leaders who perceive higher costs of delay and have sharper world views than the voters, and hence who make decisions more rapidly. We have also shown that, in electoral contests, candidates with reputations for greater decisiveness prevail and earn larger rents from holding office. These rents incentivize officials who aspire to higher office to signal decisiveness by accelerating observable decisions.

The desire to signal decisiveness induces politicians to make decisions more rapidly. However, this motive can drive politicians to make decisions far more rapidly than voters would prefer. Indeed, the signaling incentive to make fast decisions is strongest where the problem of politician indecision is least severe, which suggests that the signaling motive is not particularly well-suited to solving the politician-voter agency problem.

In settings with heterogeneous delay aversion, signaling equilibria provide the greatest rents to the least decisive politicians. Consequently, were we to extend the model by appending the familiar citizen-candidate apparatus to endogenize candidacy, the distribution of politicians would be skewed toward those with low perceived costs of delay. Those individuals are minimally affected by signaling incentives, and consequently are regarded by voters as too hesitant. In Appendix C, which studies a version of the model in which politicians signal the sharpness of their world views, we find the opposite: the most decisive politicians receive the highest utility, and would therefore be over-represented in a model with endogenous candidacy. They are also the ones who are most affected by signaling, and consequently who are most likely to overshoot the median voter's ideal.

Finally, we have addressed the effects of the institutional context on the incentive to signal decisiveness. Increased transparency, defined according to the frequency with which the electorate can effectively gauge the speed of an office holder's responses to emerging issues, leads politicians to signal through consistent but less extreme decisiveness, rather than through intermittently extreme decisiveness. Once transparency reaches a threshold

level, the effects of signaling are unambiguously beneficial. However, increases in transparency beyond that point dampen the incentive to signal, which reduces the welfare by rendering politicians consistently too hesitant from the voters' perspective.

In settings with heterogeneous ex ante policy preferences, more decisive politicians can win elections while making smaller policy concessions, thereby imposing their own agendas on the electorate. Greater partisanship amplifies the tendency to act decisively because it makes the outcomes of elections more consequential to the candidates.

While our theory takes a specific stand on the meaning of decisiveness (one that is consistent with the literature cited in the introduction), we acknowledge that common usages of the term are somewhat vague and admit complementary interpretations. For example, decisiveness can also imply that a politician does not alter a decision once it has been made. A politician who is decisive in that sense would ignore useful information that might alter his decision and improve outcomes for voters, surely a negative feature of decision-making if the story ended there. One can, however, imagine benefits to inflexibility, for example if the leader must bargain on behalf of the voters with third parties. These additional facets of decisiveness are worth exploring in future work.

Our theory also generates a number of sharp predictions that potentially lend themselves to empirical investigation. For example, is it the case that politicians who make rapid decisions are more likely to win subsequent elections? Do decisive politicians have more freedom to tailor the policy to their idiosyncratic preferences? Do politicians make slower decisions towards the end of their careers or once they have established reputations for decisiveness? We are unaware of any work in the political economy literature that has tackled these questions.

REFERENCES

- [1] Aumann, R. (1962) "Utility theory without the completeness axiom," *Econometrica*, 30, pp. 445 - 462.
- [2] Bernhardt, D.; J. Duggan; and F. Squintani (2009) "The Case for Responsible Parties," *American Political Science Review*, 103, pp. 570 - 587.
- [3] Bernheim, B.D. (1994) "A Theory of Conformity" *Journal of Political Economy*, 102 (5), pp. 841 - 877.
- [4] Bernheim, B.D. and S. Severinov (2003), "Bequests as Signals: An Explanation for the Equal Division Puzzle" *Journal of Political Economy*, 111 (4), pp. 733 - 764.
- [5] Besley, T. (2006) *Principled Agents? The Political Economy of Good Government*, Oxford University Press: Oxford, UK.
- [6] Besley, T., and S. Coate (1997), "An economic model of representative democracy," *Quarterly Journal of Economics*, 112, 85-114.

- [7] P. Bolton and A. Faure-Grimaud (2009) "Thinking Ahead: The Decision Problem," *The Review of Economic Studies*, 76, pp. 1205 - 1238.
- [8] Cho, I. and D. Kreps (1987) "Signaling games and stable equilibria," *The Quarterly Journal of Economics*, 102, pp. 179 - 222.
- [9] Cho, I. and J. Sobel (1990) "Strategic Stability and Uniqueness in Signaling Games," *Journal of Economic Theory*, 50, pp. 381 - 413.
- [10] Conlisk, J. (1988) "Optimization Cost," *Journal of Economic Behavior and Organization*, 9, pp. 213-228.
- [11] Ergin, H. and T. Sarver (2010) "A Unique Costly Contemplation Representation," *Econometrica*, 78, pp. 1285 - 1339.
- [12] Deluga, R. J. (1998) "American presidential proactivity, charismatic leadership, and rated performance," *The Leadership Quarterly*, 9, pp. 265-292
- [13] Fiorina, M.P., and S.J. Abrams (2008), "Political Polarization in the American Public," *Annual Review of Political Science*, 11, 563-388.
- [14] Garcia, H. F. (2006) "Effective leadership response to crisis," *Strategy & Leadership*, 34(1), pp. 4-10.
- [15] Holsti, Ole R. (1990) "Crisis management," in *Psychological dimensions of war*, B. Glad (ed), Sage Publications: Thousand Oaks, pp. 116-142.
- [16] Jackson, M. and M. Morelli (2007) "Political Bias and War," *The American Economic Review*, 97 (4), pp. 1353 - 1373.
- [17] Jones, J. M. (2007) "Wanted in Next President: Honesty, Strong Leadership," Gallup.com, Gallup, Inc., 4 April 2007. Web, 10 November 2016.
- [18] Kelman, S.; R. Sanders; G. Pandit; and S. Taylor (2014) "Tell It Like It Is: Groupthink, Decisiveness, and Decision-Making Among U.S. Federal Subcabinet Executives," *mimeo*.
- [19] Kopylov, I. (2009) "Choice deferral and ambiguity aversion," *Theoretical Economics*, 4, pp. 199 - 225.
- [20] Lipman, B. (1991) "How to Decide How to Decide How to ...: Modeling Limited Rationality," *Econometrica*, 59, pp. 1105 - 1125.
- [21] Maurer, T.; K. Maher; D. Ashe; D. Mitchell; M. Hein; and J. van Hein (1993) "Leadership Perceptions in Relation to a Presidential Vote," *Journal of Applied Social Psychology*, 23 (12), pp. 959 - 979.
- [22] Ok, E.; P. Ortoleva; and G. Riella (2012) "Incomplete Preferences Under Uncertainty: Indecisiveness in Beliefs versus Tastes" *Econometrica*, 80, pp. 1791 - 1808.
- [23] Pillai, R., and J. R. Meindl (1998) "Context and Charisma: A 'Meso' Level Examination of the Relationship of Organic Structure, collectivism, and Crisis to Charismatic Leadership," *Journal of Management*, 24(5), pp. 643-671.
- [24] Pillai, R. and E. Williams (1998) "Does Leadership Matter in the Political Arena? Voter Perceptions of Candidates' Transformational and Charismatic Leadership and the 1996 U.S. Presidential Vote," *Leadership Quarterly*, 9 (3), pp. 397 - 416.
- [25] Pillai, R.; E. Williams; K. Lowe; and D. Jung (2003) "Personality, transformational leadership, trust, and the 2000 U.S. presidential vote," *Leadership Quarterly*, 14, pp. 161 - 192.
- [26] Rapoport, R.; K. Metcalf; and J. Hartman (1989) "Candidate Traits and Voter Inferences: An Experimental Study," *The Journal of Politics*, 51 (4) , pp. 917 - 932.

- [27] Shabad, G. and K. Andersen (1979) "Candidate Evaluations by Men and Women," *Public Opinion Quarterly*, pp. 18 - 35.
- [28] Shamir, B. (1994) "Ideological Position, Leaders' Charisma, and Voting Preferences: Personal Vs. Partisan Elections," *Political Behavior*, 16 (2), pp. 265 - 287.
- [29] Simon, A.F. (2006) "Leadership and managing ambivalence," *Consulting Psychology Journal: Practice and Research*, 58, pp. 91–105
- [30] Simpson, P., R. French, C.E. Harvey (2002) "Leadership and negative capability," *Human Relations*, 55, pp. 1209–1226.
- [31] Williams, E.; R. Pillai.; K. Lowe; D. Jung; and D. Herst (2009) "Personality, transformational leadership, trust, and the 2000 U.S. presidential vote," *Leadership Quarterly*, 20, pp. 70 - 86.
- [32] Yukl, G. (1999) "An evaluations of conceptual weaknesses in transformational and charismatic leadership theories," *Leadership Quarterly*, 10(2), 285-305.

APPENDIX A. PROOFS

Proposition 1. *From the perspective of voter i , the ideal politician has minimal uncertainty about his own preferences ($\beta_P = \underline{\beta}$) and perceives the cost of delay per unit time to be*

$$c_P = \begin{cases} c_i \frac{\alpha + \underline{\beta}}{\alpha - \underline{\beta}} & \text{if } \alpha > \underline{\beta} \text{ and the result is less than } \bar{c} \\ \bar{c} & \text{otherwise} \end{cases} .$$

Proof. Initially let us assume $\alpha > \beta_P$. The decision time a voter would like the politician to choose can be derived from the following first order condition, where τ^v denotes the voter's preferred time for the politician's decision

$$\begin{aligned} \frac{\partial}{\partial \tau} W_i^P(\tau) &= -\frac{\partial}{\partial \tau} \left[\frac{\alpha}{\tau + \phi} + \frac{\beta_P}{\phi} \frac{\tau}{\tau + \phi} + c_i \tau \right] \\ &= \frac{\alpha}{(\tau^v + \phi)^2} + \frac{\beta_P}{\phi} \left(\frac{\tau^v}{(\tau^v + \phi)^2} - \frac{1}{\tau^v + \phi} \right) - c_i \\ &= \frac{\alpha - \beta_P}{(\tau^v + \phi)^2} - c_i = 0. \end{aligned}$$

Simplifying this we find

$$(15) \quad \tau^v = \sqrt{\frac{\alpha - \beta_P}{c_i}} - \phi.$$

The second order condition is

$$\frac{\partial^2}{\partial \tau^2} W_i^P(\tau) = \frac{\partial}{\partial \tau} \left[\frac{\alpha - \beta_P}{(\tau + \phi)^2} - c_i \right] < 0$$

which implies the voter's utility is concave in τ .

Comparing equation 15 with equation 3 we find

$$\tau^* \left(\beta_P, c_i \frac{\alpha + \beta_P}{\alpha - \beta_P} \right) = \sqrt{\frac{\alpha + \beta_P}{c_i} \left(\frac{\alpha - \beta_P}{\alpha + \beta_P} \right)} - \phi = \tau^v.$$

In other words, this cost induces the politician to make decisions at the voter's preferred pace. If $c_i \frac{\alpha + \beta_P}{\alpha - \beta_P} > \bar{c}$, the voter's most preferred politician will have the highest cost possible, \bar{c} .

Now consider the case where $\alpha \leq \beta_P$, in which case $\frac{\partial}{\partial \tau} W_i^P(\tau) < 0$ for all τ . Again, this implies that voter i would like the politician to make decisions as quickly as possible, which implies that voter i would like the politician's information gathering to be as high as possible (i.e., equal to \bar{c}) \square

Our proofs for the slope and global incentive compatibility for equilibria of the game where politicians signal aversion to delay closely are based on the ODE that defines c as a function of τ

$$(16) \quad \frac{\partial c^S}{\partial \tau} = \lambda c \left[\frac{\partial \Pi(c^S(\tau), \hat{c})}{\partial \hat{c}} \Big|_{\hat{c}=c^S(\tau)} \right]^{-1} \left[1 - \left(\frac{\tau^*(c^S(\tau)) + \phi}{\tau + \phi} \right)^2 \right].$$

The following lemma insures that the right-hand side of equation 16 is well defined.

Lemma 3. *There exist finite $d_L, d_U > 0$ such that $d_L < \frac{\partial \Pi(c, \hat{c})}{\partial \hat{c}} \Big|_{\hat{c}=c} < d_U$ for all $c \in [\underline{c}, \bar{c}]$.*

Proof. Taking the derivative, we have

$$(17) \quad \frac{\partial \Pi(c, \hat{c})}{\partial \hat{c}} \Big|_{\hat{c}=c} = 2\beta \left[\frac{1}{\phi} - \frac{1}{\tau^*(c) + \phi} \right] f_c(c).$$

Since $\tau^*(c) \geq \tau^*(\bar{c}) > 0$, our claim must hold. \square

Proposition 2. $c^S(\tau)$ is strictly decreasing in τ with $\tau < \tau^*(c^S(\tau))$ for $\tau < \tau^*(\underline{c})$ and strictly increasing in τ with $\tau > \tau^*(c^S(\tau))$ for $\tau > \tau^*(\underline{c})$.

Proof. Substituting the initial condition into equation 16, we obtain $\frac{\partial c^S}{\partial \tau} \Big|_{\tau=\tau^*(\bar{c})} = 0$. Using equation 16 along with lemma 3, we see that $\frac{\partial c^S}{\partial \tau} > 0$ when $\tau > \tau^*(c^S(\tau))$, and $\frac{\partial c^S}{\partial \tau} < 0$ when $\tau < \tau^*(c^S(\tau))$.

First consider $\tau > \tau^*(\underline{c})$. Starting at $\tau^*(\underline{c})$, a small increase in τ has a negligible effect on c^S , leaving us with $\tau > \tau^*(c^S(\tau))$, and hence $\frac{\partial c^S}{\partial \tau} > 0$. As τ increases, the slope remains strictly positive unless we reach a point at which $\tau = \tau^*(c^S(\tau))$. But that is impossible, because $\tau^*(c) < \tau^*(\underline{c})$ for all $c > \underline{c}$.

Next consider $\tau < \tau^*(\underline{c})$. Starting at $\tau^*(\underline{c})$, a small decrease in τ has a negligible effect on c^S , leaving us with $\tau < \tau^*(c^S(\tau))$, and hence $\frac{\partial c^S}{\partial \tau} < 0$. As τ decreases, the slope remains strictly negative unless we reach a value τ' at which $\tau' = \tau^*(c^S(\tau'))$. But that is impossible, because (i) $\frac{\partial \tau^*}{\partial c}$ is strictly negative and bounded away from 0 on $[\underline{c}, \bar{c}]$, and (ii) from equation 16 and lemma 3, $\frac{\partial c^S}{\partial \tau}$ would converge to zero as $(\tau, c^S(\tau))$ converged to $(\tau', \tau^*(c^S(\tau')))$; thus, further reductions in τ would widen the gap between $c^S(\tau)$ and $(\tau^*)^{-1}(\tau)$. \square

Proposition 3. *There exists λ^* such that for $\lambda > \lambda^*$, τ^1 is globally incentive-compatible. τ^2 is never globally incentive compatible.*

Proof. We begin by showing that τ^1 converges to τ^* as $\lambda \rightarrow \infty$, and establishing a lower bound on $\frac{\partial \tau^1}{\partial c}$. To this end, we define $R(\lambda) = \max_{c \in [\underline{c}, \bar{c}]} \frac{\tau^*(c) + \phi}{\tau^1(c) + \phi} > 1$. Continuity of τ^* and τ^1 ensures existence of the maximum.

Lemma 4. *For any λ and all $c \in [\underline{c}, \bar{c}]$, we have $\frac{\partial \tau^1}{\partial c} < \frac{-d_L}{\lambda c(R(\lambda)^2 - 1)}$. Furthermore, $\lim_{\lambda \rightarrow \infty} R(\lambda) = 1$.*

Proof. Fix a value of λ . From equation 10 and lemma 3, we have $d_L < \lambda c(1 - R(\lambda)^2) \frac{\partial \tau^1}{\partial c}$, or equivalently $\frac{\partial \tau^1}{\partial c} < \frac{-d_L}{\lambda c(R(\lambda)^2 - 1)}$, for all $c \in [\underline{c}, \bar{c}]$, which establishes the first part of the lemma.

For the second part of the lemma, we begin by observing that $\frac{\partial \tau^*}{\partial c} = \frac{-\sqrt{\alpha + \beta}}{2c^{3/2}}$, which is bounded above by $\frac{-g}{c}$ where $g \equiv \frac{1}{2} \sqrt{\frac{\alpha + \beta}{c}}$. Now fix $r > 1$, and define $\lambda_r \equiv r \frac{d_H}{g(r^2 - 1)}$. From equation 10 and the definitions of d_H and g , it follows that, if $\lambda > \lambda_r$ and $\frac{\tau^*(c) + \phi}{\tau^1(c) + \phi} \geq r$, we have $0 > r \frac{\partial \tau^1}{\partial c} > \frac{\partial \tau^*}{\partial c}$.

Next we show that, for all $\lambda > \lambda_r$, we have $R(\lambda) < r$. Suppose on the contrary that, for such λ , there is some c for which $\frac{\tau^*(c) + \phi}{\tau^1(c) + \phi} \geq r$. Let c' be the smallest value for which $\frac{\tau^*(c') + \phi}{\tau^1(c') + \phi} = r$. Differentiating, we obtain

$$\begin{aligned} \frac{d}{dc} \left(\frac{\tau^*(c) + \phi}{\tau^1(c) + \phi} \right) \Big|_{\beta = \beta'} &= \frac{\frac{d\tau^*}{dc} (\tau^1(c') + \phi) - \frac{d\tau^1}{dc} (\tau^*(c') + \phi)}{(\tau^1(c') + \phi)^2} \\ &= \frac{\frac{d\tau^*}{dc} - \frac{d\tau^1}{dc} r}{\tau^1(c') + \phi} \\ &< 0. \end{aligned}$$

It follows that a small decrease in c from c' would result in $\frac{\tau^*(c) + \phi}{\tau^1(c) + \phi} > r$. But we know that $\frac{\tau^*(\underline{c}) + \phi}{\tau^1(\underline{c}) + \phi} = 1 < r$. Consequently, there would have to be some $c'' \in (\underline{c}, c')$ for which

$\frac{\tau^*(c'')+\phi}{\tau^1(c'')+\phi} = r$. But that contradicts the definition of c' . We conclude that, for any $r > 1$, we have $R(\lambda) < r$ for all $\lambda > \lambda_r$. The lemma follows directly. \square

Lemma 5. *There exists λ^* such that for all $\lambda > \lambda^*$, we have $\frac{\partial^2 V(c, \hat{c})}{\partial c \partial \hat{c}} > 0$ for all $c \in [\underline{c}, \bar{c}]$.*

Proof. Using equation 3 to substitute for $\tau^*(c)$ in equation 8 and differentiating, we obtain

$$(18) \quad \begin{aligned} \frac{\partial^2 V(c, \hat{c})}{\partial \hat{c} \partial c} &= (\tau^*(\hat{c}) - \tau^*(c)) f_c(\hat{c}) - \lambda \frac{\partial \tau^S(\hat{c})}{\partial \hat{c}} \\ &> (\tau^*(\bar{c}) - \tau^*(\underline{c})) \bar{f}_c + \frac{d_L}{c(R(\lambda)^2 - 1)} \end{aligned}$$

where \bar{f}_c is an upper bound on the density of c , and the inequality makes use of lemma 4. While the first term is negative, lemma 4 implies the second term becomes unboundedly large and positive as $\lambda \rightarrow \infty$. Accordingly, the entire expression is strictly positive for λ sufficiently large. \square

Now we prove the first part of the proposition. In light of the fact that $\frac{\partial^2 V(c, \hat{c})}{\partial c \partial \hat{c}} > 0$, the first-order condition for claiming to be of type \hat{c} , $\frac{\partial V(c, \hat{c})}{\partial \hat{c}} = 0$, can be satisfied by at most one type, which is by construction $c = \hat{c}$. Consequently the optimal value of \hat{c} for each c must be c, \underline{c} , or \bar{c} . We can rule out \underline{c} on the grounds that $\left. \frac{\partial V(c, \hat{c})}{\partial \hat{c}} \right|_{c, \hat{c}=\underline{c}} = 0$ and $\frac{\partial^2 V(c, \hat{c})}{\partial c \partial \hat{c}} > 0$; similarly for \bar{c} . We conclude that τ^1 is globally incentive-compatible.

Turning to the second part of the proposition, we know that $\lim_{c \rightarrow \underline{c}} \frac{\partial \tau^2}{\partial c} = +\infty$. Accordingly, the analog of equation 18 tells us that $\frac{\partial^2 V(c, \hat{c})}{\partial c \partial \hat{c}} < 0$ for c and \hat{c} in a neighborhood of \underline{c} . But in that case, local incentive compatibility is violated: for c_1 and c_2 close to \underline{c} , the fact that $V(c_1, c_1) > V(c_1, c_2)$ would imply $V(c_2, c_1) > V(c_2, c_2)$. \square

Proposition 4. *$V(c, c)$ is decreasing in c .*

Proof. Consider $c > c'$ and use the following decomposition of $V(c, c) - V(c', c')$:

$$V(c, c) - V(c', c') = [V(c, c) - V(c', c)] + [V(c', c) - V(c', c')].$$

The second term is weakly negative since incentive compatibility requires $V(c', c') \geq V(c', c)$. Now we turn to the first term, which represent the utility difference between agents of type c and \hat{c} when both claim to be of type c .

$$(19) \quad \begin{aligned} V(c, c) - V(c', c) &= - \int_{\underline{c}}^c \left[\frac{\alpha + \beta}{\tau^*(c) + \phi} + c\tau^*(c) - \frac{\alpha + \beta}{\tau^*(c') + \phi} - c'\tau^*(c') \right] f_c(s) ds \\ &\quad - \int_c^{\bar{c}} (c - c') \tau^*(s) f_c(s) ds - \lambda (c - c') \tau^S(c). \end{aligned}$$

While the second and third terms are clearly negative since $c > c'$, we need a bit more work to sign the first term. Note that

$$(20) \quad \frac{\alpha + \beta}{\tau^*(\tilde{c}) + \phi} + \tilde{c}\tau^*(\tilde{c}) = 2\sqrt{\tilde{c}(\alpha + \beta)} - \tilde{c}\phi.$$

The first term in equation 19 is negative if the expression on the right-hand side of equation 20 is increasing in \tilde{c} . Notice that

$$\frac{d}{dc} \left[2\sqrt{\alpha + \beta}\sqrt{c} - c\phi \right] = \sqrt{\frac{\alpha + \beta}{c}} - \phi = \tau^*(c) > 0.$$

Therefore $V(c, c) < V(c', c)$ and, as a result, $V(c, c) < V(c', c')$. \square

Proposition 5. *The symmetric equilibrium maximizes the payoff for every type of politician within the set of fully separating equilibria.*

Proof. Suppose we have a separating equilibrium with action functions $\tau^S(c) = (\tau_1^S(c), \dots, \tau_m^S(c))$. Defining

$$(21) \quad \Gamma(c, \tau) \equiv \sum_{m=1}^M \left(\frac{\alpha + \beta}{\tau_m + \phi} + c\tau_m \right),$$

we can rewrite the first-order condition for type c 's optimal choice as

$$(22) \quad \left. \frac{\partial \Pi(c, \hat{c})}{\partial \hat{c}} \right|_{\hat{c}=c} = \frac{\partial \Gamma(c, \tau^S(c))}{\partial \tau} \frac{d\tau^S(c)}{dc},$$

where $\frac{\partial \Gamma(c, \tau^S(c))}{\partial \tau}$ and $\frac{d\tau^S(c)}{dc}$ are M -dimensional vectors.³¹ We are interested in determining type c 's total payoff in equilibrium. Definitionally,

$$(23) \quad V(c, c) = \Pi(c, c) - \Gamma(c, \tau^S(c)),$$

and using the Envelope Theorem we have

$$\frac{dV(c, c)}{dc} = \frac{d\Pi(c, c)}{dc} - \frac{\partial \Gamma(c, \tau^S(c))}{\partial c}.$$

Notice that only the final term depends on the particular separating equilibrium. Let τ^0 denote the symmetric separating equilibrium with payoffs V^0 , and τ^A denote an asymmetric

³¹We assume for convenience that each τ_m^S is differentiable. However, the argument only requires differentiability of $\Gamma(c, \tau^S(c))$, which is a slightly weaker condition. As an example, consider $M = 1$ with an equilibrium where τ_1^S and τ_2^S are differentiable. We could define another equilibrium $\tilde{\tau}_1^S$ and $\tilde{\tau}_2^S$ where

$$\tilde{\tau}_1^S(c) = \begin{cases} \tau_1^S(c) & \text{if } c \text{ is a rational number} \\ \tau_2^S(c) & \text{otherwise} \end{cases}$$

and $\tilde{\tau}_2^S(c)$ is defined symmetrically. Obviously the equilibrium strategies would not be differentiable.

separating equilibrium with payoffs V^A . To demonstrate that payoffs in the symmetric separating equilibrium are strictly higher than in the asymmetric separating equilibrium, we will establish the following Property (capitalized for clarity of subsequent references): if it were the case for some c that either (i) $V^0(c, c) = V^A(c, c)$ and $\tau^0(c) \neq \tau^A(c)$, or (ii) $V^0(c, c) < V^A(c, c)$, then we would have $\frac{dV^0(c, c)}{dc} > \frac{dV^A(c, c)}{dc}$.³² To understand why this property delivers the desired conclusion, note that $V^A(c', c') - V^0(c', c')$ would then shrink with c' over $[\underline{c}, c]$, violating the boundary conditions $V^0(\bar{c}, \bar{c}) = V^A(\bar{c}, \bar{c}) = \Pi(\bar{c}, \bar{c}) - \Gamma(c, \tau^*(\bar{c}))$. In light of equation 23, we can rewrite the Property as follows: if it were the case for some c that either (i)' $\Gamma(c, \tau^0(c)) = \Gamma(c, \tau^A(c))$ and $\tau^0(c) \neq \tau^A(c)$, or (ii)' $\Gamma(c, \tau^0(c)) > \Gamma(c, \tau^A(c))$, then we would have $\frac{\partial \Gamma(c, \tau^0(c))}{\partial c} < \frac{\partial \Gamma(c, \tau^A(c))}{\partial c}$.

We now establish the Property. Supposing condition (i)' were satisfied for some $c > \underline{c}$, we would begin by defining³³

$$\bar{\tau}_m = \begin{cases} \tau_m^A(c) & \text{if } \tau_m^A(c) \leq \tau^*(c) \\ \tau \leq \tau^*(c) \text{ s.t. } \frac{\alpha+\beta}{\tau+\phi} + c\tau = \frac{\alpha+\beta}{\tau_m^A(c)+\phi} + c\tau_m^A(c) & \text{otherwise} \end{cases}.$$

Let $Q \equiv \{m \mid \tau_m^A(c) > \tau^*(c)\}$. Then

$$\frac{\partial \Gamma(c, \bar{\tau})}{\partial c} - \frac{\partial \Gamma(c, \tau^A(c))}{\partial c} = \sum_{m \in Q} (\bar{\tau}_m - \tau^A(c)) \leq 0,$$

with strict inequality if Q is non-empty.

If $\tau^0(c) = \bar{\tau}$, we are done. If not, then since $\Gamma(c, \tau^A(c)) = \Gamma(c, \bar{\tau})$ by construction, there must exist i and j such that $\bar{\tau}_i > \tau^0(c) > \bar{\tau}_j$.

Define the function $\tilde{\tau}(\tau_i)$ as follows: $\tilde{\tau}_i(\tau_i) = \bar{\tau}_i$, $\tilde{\tau}_k(\tau_i) = \bar{\tau}_k$ for $k \neq i, j$, and $\Gamma(c, \tilde{\tau}(\tau_i)) = \Gamma(c, \bar{\tau})$. In other words, $\tilde{\tau}_j(\tau_i)$ indicates how τ_j must vary in response to changes in τ_i to keep the value of Γ constant at its equilibrium value. Implicit differentiation reveals that for $\bar{\tau}_i > \bar{\tau}_j$

$$\left. \frac{d\tilde{\tau}_j}{d\tau_i} \right|_{\tau_i = \bar{\tau}_i} = - \frac{\frac{\alpha+\beta}{(\bar{\tau}_i+\phi)^2} - c}{\frac{\alpha+\beta}{(\tilde{\tau}_j(\tau_i)+\phi)^2} - c}.$$

³²Suppose our claim is true. Then if either condition (i) or (ii) holds for c , then condition (ii) must hold for all $c' \in (\underline{c}, c)$.

³³Although we ruled out $\tau_m^A(c) > \tau^*(c)$ when $M = 1$, in principle this need be true when $M > 1$. $\bar{\tau}_m$ is defined so that an equivalent signaling cost is incurred, but all of the actions are in the intuitive direction of $\bar{\tau}_m(c) \leq \tau^*(c)$.

Plainly, there exists a unique value $\tau_i^e < \tau^*(c)$ such that $\tilde{\tau}_j(\tau_i^e) = \tau_i^e$. For $\tau_i \in [\tau_i^e, \tau_i^A(c)]$, we have³⁴

$$\begin{aligned} \frac{d}{d\tau_i} \left(\frac{\partial \Gamma(c, \tilde{\tau}(\tau_i))}{\partial c} \right) &= \frac{d}{d\tau_i} (\tau_i + \tilde{\tau}_j(\tau_i)) \\ &= 1 + \frac{d\tilde{\tau}_j}{d\tau_i} > 0, \end{aligned}$$

where the final inequality follows from the fact that $\tilde{\tau}_j(\tau_i) < \tau_i$ (and hence $\frac{d\tilde{\tau}_j}{d\tau_i} > -1$). It follows that $\frac{\partial \Gamma(c, \tilde{\tau}(\tau_i^e))}{\partial c} < \frac{\partial \Gamma(c, \bar{\tau})}{\partial c}$ since $\bar{\tau}_i > \tau^0(c)$ is being reduced in this equalization step. Through repeated application of this equalization argument, we conclude that $\frac{\partial \Gamma(c, \tau^0(c))}{\partial c} < \frac{\partial \Gamma(c, \bar{\tau})}{\partial c} \leq \frac{\partial \Gamma(c, \tau^A(c))}{\partial c}$, as desired.

Next, supposing condition (ii)' were satisfied for some $c > \underline{c}$, we would begin by defining τ' s.t. $\tau'_1 = \tau'_2 = \dots = \tau'_M < \tau^*(c)$ and $\Gamma(c, \tau') = \Gamma(c, \tau^A(c))$. By the same argument as for condition (ii)', we infer $\frac{\partial \Gamma(c, \tau')}{\partial c} \leq \frac{\partial \Gamma(c, \tau^A(c))}{\partial c}$. (The inequality is weak because we include the possibility that $\tau' = \tau^A(c)$). Because $\Gamma(c, \tau^0(c)) < \Gamma(c, \tau^A(c)) = \Gamma(c, \tau')$ by assumption, we have $\tau_m^0(c) < \tau'_m$. Accordingly,

$$\frac{\partial \Gamma(c, \tau^0(c))}{\partial c} - \frac{\partial \Gamma(c, \tau')}{\partial c} = \sum_{m=1}^M (\tau_m^0(c) - \tau'_m) < 0.$$

It follows that $\frac{\partial \Gamma(c, \tau^0(c))}{\partial c} < \frac{\partial \Gamma(c, \tau^A(c))}{\partial c}$, as desired.

Having established that the Property holds, the Proposition follows for the reasons given above. \square

Proposition 6. $\tau^0(c)$ converges to $\tau^*(c)$ uniformly as $M \rightarrow \infty$.

Proof. A comparison between Equations 11 and 10 reveals that the two are isomorphic, inasmuch as one can simply absorb M into λ . The Proposition then follows directly from Lemma 4. \square

Lemma 1. *If a voter with characteristics (μ_M, c_M) strictly prefers candidate i to candidate j , then so does a majority of voters.*

Proof. The difference in utility that voter (μ_i, c_i) receives if candidate 1 is elected versus candidate 2 where candidate j commits to a power sharing agreement δ_j is

$$\Delta_i = 2\mu_i(2\mu + \delta_1 - \delta_2) - (\tau_1 - \tau_2)c_i + \left[(\delta_2 - \mu)^2 - (\delta_1 + \mu)^2 + \left(\frac{\alpha - \beta_2}{\tau_2 + \phi} - \frac{\alpha - \beta_1}{\tau_1 + \phi} \right) + \frac{\beta_2 - \beta_1}{\phi} \right],$$

³⁴Since $\bar{\tau}_i > \bar{\tau}_j$, it must be that $\bar{\tau}_i > \tau_i^e > \bar{\tau}_j$

where the bracketed term is independent of the voter's type. Clearly the indifference curves implied by Δ_i are linear, which implies that the voters that prefer candidate 1 to candidate 2 can be described using a half-plane. The claim regarding the pivotality of (μ_M, c_M) follows from elementary geometric arguments invoking the symmetry of the distribution of types in each \square

Lemma 2. $\frac{\partial}{\partial \hat{c}_1} \Phi_M(\hat{c}_1, \hat{c}_2)$ has the same sign as $c_M \frac{\alpha+\beta}{\alpha-\beta} - \hat{c}_1$, and $\frac{\partial}{\partial \hat{c}_2} \Phi_M(\hat{c}_1, \hat{c}_2)$ has the same sign as $\hat{c}_2 - c_M \frac{\alpha+\beta}{\alpha-\beta}$.

Proof. Suppose the ideal politician type for the median voter has an aversion to delay larger than \bar{c} . Since the median agents prefers politicians that have higher aversion to delay, an increase in the perceived aversion of delay in candidate 1 increases the appeal of that candidate (i.e., Φ_M rises). An identical argument implies $\frac{\partial}{\partial \hat{c}_2} \Phi_M(\hat{c}_1, \hat{c}_2) < 0$. If the median voter's ideal politician has a cost within $[\underline{c}, \bar{c}]$, then the sign convention in our lemma captures the single-peakedness of the median voter's ideal politician type. \square

Proposition 7. *The Nash equilibria depend on the model's parameters as follows:*

- (1) If $\mu^2 \geq \Phi_M > 0$, the unique pure strategy Nash equilibrium involves $\delta_1 = \sqrt{\Phi_M} - \mu$, $\delta_2 = \mu$, and the election of candidate 1.
- (2) If $\Phi_M > \mu^2$, all pure strategy Nash equilibria involve $\delta_1 = 0$ and the election of candidate 1.
- (3) If $\mu^2 \geq -\Phi_M > 0$, the unique pure strategy Nash equilibrium involves $\delta_2 = \mu - \sqrt{-\Phi_M}$, $\delta_1 = -\mu$, and the election of candidate 2.
- (4) If $-\Phi_M > \mu^2$, all pure strategy Nash equilibria involve $\delta_2 = 0$ and the election of candidate 2.
- (5) If $\Phi_M = 0$, then $\delta_1 = \delta_2 = -\mu$ and each candidate is elected with probability $\frac{1}{2}$.

Proof. In each case, it is easily verified that the indicated actions constitute a Nash equilibrium. In cases (ii) and (iv), any values for δ_1 and δ_2 , respectively, will suffice.

We establish uniqueness for case (i) by dividing the alternatives into the following categories.

- First suppose $\delta_1 = \sqrt{\Phi_M} - \mu$ and $\delta_2 \neq \mu$, in which case candidate 1 wins the election. Then there exists $\varepsilon > 0$ such that candidate 1 can deviate to $\tilde{\delta}_1 = \sqrt{\Phi_M} - \mu + \varepsilon$, still win the election, and achieve a preferred outcome.
- Next suppose $\delta_1 < \sqrt{\Phi_M} - \mu$ and candidate 1 wins. Then candidate 1 could deviate to $\tilde{\delta}_1 = \sqrt{\Phi_M} - \mu$, still win the election, and achieve a preferred outcome.
- Next suppose $\delta_1 < \sqrt{\Phi_M} - \mu$ and candidate 1 loses. Then candidate 1 could deviate to $\tilde{\delta}_1 = \sqrt{\Phi_M} - \mu$, win the election, and achieve a preferred outcome.

- Next suppose $\delta_1 > \sqrt{\Phi_M} - \mu$ and candidate 1 wins. Then candidate 2 could deviate to $\tilde{\delta}_2 = \mu$, win the election, and achieve a preferred outcome.
- Finally suppose $\delta_1 > \sqrt{\Phi_M} - \mu$ and candidate 1 loses. Then candidate 1 could deviate to $\tilde{\delta}_1 = \sqrt{\Phi_M} - \mu$, win the election, and achieve a preferred outcome.

The argument for uniqueness is symmetric in case (iii), and proceeds similarly in case (v). \square

Proposition 8. *In settings where they signal aversion to delay, $\tau^S(c)$ is decreasing in μ .*

Proof. Consider $\mu > \tilde{\mu}$, and suppose contrary to the proposition that there exists some $c' > \underline{c}$ such that $\tau^S(c'; \mu) \geq \tau^S(c'; \tilde{\mu})$. Then there must be some $c \in [\underline{c}, c')$ with $\tau^S(c; \mu) \geq \tau^S(c; \tilde{\mu})$ such that $\frac{d\tau^S(c; \mu)}{dc} \leq \frac{d\tau^S(c; \tilde{\mu})}{dc}$ – otherwise we would have $\tau^S(c; \mu) > \tau^S(c; \tilde{\mu})$, which violates the boundary condition.³⁵ However,

$$\begin{aligned} \left. \frac{\partial \Pi(c, \hat{c})}{\partial \hat{c}} \right|_{\hat{c}=c} &= \lambda c \left[1 - \left(\frac{\tau^*(c) + \phi}{\tau^S(c) + \phi} \right)^2 \right] \frac{\partial \tau^S}{\partial \hat{c}} \\ \frac{d\tau^S(c; \mu)}{dc} &= \frac{1}{\lambda c} \left[1 - \left(\frac{\tau^*(c) + \phi}{\tau^S(c; \mu) + \phi} \right)^2 \right]^{-1} \left(\left. \frac{\partial \Pi(c, \hat{c}; \mu)}{\partial \hat{c}} \right|_{\hat{c}=c} \right) \\ &> \frac{1}{\lambda c} \left[1 - \left(\frac{\tau^*(c) + \phi}{\tau^S(c; \tilde{\mu}) + \phi} \right)^2 \right]^{-1} \left(\left. -\frac{\partial \Pi(c, \hat{c}; \mu)}{\partial \hat{c}} \right|_{\hat{c}=c} \right) \\ &> \frac{1}{\lambda c} \left[1 - \left(\frac{\tau^*(c) + \phi}{\tau^S(c; \tilde{\mu}) + \phi} \right)^2 \right]^{-1} \left(\left. -\frac{\partial \Pi(c, \hat{c}; \tilde{\mu})}{\partial \hat{c}} \right|_{\hat{c}=c} \right) \\ &= \frac{d\tau^S(c; \tilde{\mu})}{dc}, \end{aligned}$$

where the first inequality follows from $\tau^S(c; \mu) \geq \tau^S(c; \tilde{\mu})$, and the second from $\frac{\partial}{\partial \mu} \frac{\partial \Pi}{\partial \hat{c}} > 0$ (which we demonstrated in the text). Thus we have a contradiction. \square

APPENDIX B. AN EXPLICIT MODEL OF FUZZY PREFERENCES

In this appendix, we provide explicit microfoundations for our interpretation of uncertainty about x_i as “fuzzy preferences” and/or a “weak moral compass.”

³⁵If $\tau^S(c'; \mu) > \tau^S(c'; \tilde{\mu})$, then the existence of a c such that $\tau^S(c; \mu) > \tau^S(c; \tilde{\mu})$ follows from the continuity of $\tau^S(c; \mu)$ and $\tau^S(c; \tilde{\mu})$. If $\tau^S(c'; \mu) = \tau^S(c'; \tilde{\mu})$, then we know there exists a c such that $\tau^S(c; \mu) > \tau^S(c; \tilde{\mu})$ because $\frac{\partial \Pi}{\partial c \partial \mu} > 0$, which implies

$$\left. \frac{\partial \tau^S(c; \mu)}{\partial c} \right|_{c=c'} > \left. \frac{\partial \tau^S(c; \tilde{\mu})}{\partial c} \right|_{c=c'}$$

Suppose the outcome of policy p is $(K + 1)$ -dimensional. Label these dimensions $k = 0, 1, \dots, K$. Also assume that disagreements among citizens arise from the fact that they weigh these dimensions differently. Everyone agrees that z_k would be the ideal policy for someone who places all weight on outcome k , and that the loss function would be $(p - z_k)^2$. However, they disagree about the relative importance of the different dimensions. Each citizen i potentially cares about all of these dimensions and would like to minimize a weighted average of the loss functions:

$$\sum_{k=0}^K \omega_{ik} (p - z_k)^2$$

where $\sum_{k=0}^K \omega_{ik} = 1$.

To model “fuzzy” preferences, we allow for the possibility that citizens are initially uncertain about the importance they will ultimately attach to the various dimensions of the policy’s outcomes. Consequently, they view the ω_{ik} as random variables. Over time, they may acquire information that leads them to become more certain about the weights they will ultimately employ. For example, a politician may learn about the considerations her donors and/or constituents consider important, or she may clarify her own values through internal reflection.

In this setting, an individual with sharp preferences (alternatively, a strong moral compass) is one with little or no uncertainty about the weights, ω_{ik} : she knows what is important to her and is not easily convinced to change her priorities. In contrast, an individual with fuzzy preferences (alternatively, a weak moral compass) is one with substantial uncertainty about the weights: she has no deep convictions about what is important and is easily influenced.

Observe that we can rewrite the objective function as:

$$\begin{aligned} \sum_{k=0}^K \omega_{ik} (p - z_k)^2 &= p^2 - \sum_{k=0}^K 2p\omega_{ik}z_k + \sum_{k=0}^K \omega_{ik}z_k^2 \\ &= \left[p^2 - \sum_{k=0}^K 2p\omega_{ik}z_k + \sum_{j=0}^K \sum_{k=0}^K \omega_{ik}z_jz_k \right] + \sum_{k=0}^K \omega_{ik}z_k^2 - \sum_{j=0}^K \sum_{k=0}^K \omega_{ik}z_jz_k \\ &= \left(p - \sum_{k=0}^K \omega_{ik}z_k \right)^2 + F(z_0, \dots, z_K, \omega_{i0}, \dots, \omega_{iK}). \end{aligned}$$

Policy preferences plainly do not depend on the function F , so we can ignore it. Let $\bar{\omega}_k$ denote the population mean of ω_{ik} . Define:

$$\begin{aligned}\theta &= \sum_{k=0}^K \bar{\omega}_k z_k \\ x_i &= \sum_{k=0}^K (\omega_{ik} - \bar{\omega}_k) z_k.\end{aligned}$$

Then citizen i 's preferences over policies are governed by the loss function $(p - \theta - x_i)^2$.

Our analysis also assumes that people are uncertain about objective factors that bear on the identity of the optimal policy. We simplify by keeping these two sources of uncertainty separate. In particular, assume that ex ante uncertainty is confined to z_0 and ω_{ik} for $k \geq 1$, and that all citizens attach the same weight to dimension 0 (i.e., $\omega_{i0} = \omega_0$ for all i).³⁶ While this assumption is obviously special, it proves analytically convenient, because it guarantees that θ and x_i are independently distributed random variables, exactly as assumed in the text. However, the same conceptual principles apply when all z_k and ω_{ik} are random, inasmuch as the resolution of a politician's welfare weights introduces gratuitous policy variation from the perspective of the voters.

APPENDIX C. ONLINE APPENDIX: SIGNALING β

For the purpose of this section, we assume all agents perceive the same costs of delay, c . Candidates therefore differ only in the precision of their knowledge of their idiosyncratic preferences, β_P . As in the case where politicians signal an aversion to delay, the victor in the stage 2 election will be the candidate voters believe has the greatest decisiveness – in other words, the one with the lowest value of $\hat{\beta}$. Since low values of β_P yield rapid decisions in stage 3, each candidate has an incentive to signal decisiveness in stage 1. Throughout this extension we include ideological adjustments as per Section 6 to illustrate how these adjustments affect our proofs.

Now consider the stage 1 decision problem facing a politician with type β who expects to play the role of candidate 1 in the stage 2 election. (The analysis for a candidate who expects to play the role of candidate 2 is symmetric.) From the preceding discussion, the candidate knows he will win the election if voters believe he has sharper views about how to weigh different aspects of the problem than his opponent. Because we are studying fully separating equilibria, the stage 1 outcome will fully reveal his opponent's type. Therefore,

³⁶The analysis trivially generalizes to the case in which uncertainty is confined to either z_k or ω_k for each k . For our purposes, the key simplifying assumption is that the two sources of uncertainty are orthogonal.

when choosing the image he wishes to project in stage 1 ($\widehat{\beta}$), he knows he will win if it turns out that his chosen $\widehat{\beta}$ is less than the actual value of his opponent's β parameter, and lose if it is greater. For the purpose of simplifying some of the analytic expressions, we assume that μ is large enough that δ is strictly less than μ on the entire support of f_β (i.e., $\Phi_M(\underline{\beta}, \overline{\beta}) < \mu$). Using equations 1 and 5 along with proposition 7, we can write the stage 3 payoff of a politician with type β who chooses a perception $\widehat{\beta}$ as follows:

$$(24) \quad \Pi(\beta, \widehat{\beta}) = - \int_{\widehat{\beta}}^{\overline{\beta}} \left[\left(\mu - \sqrt{\Phi_M(\widehat{\beta}, s)} \right)^2 + \frac{\alpha + \beta}{\tau^*(\beta) + \phi} + c\tau^*(\beta) \right] f_\beta(s) ds \\ - \int_{\underline{\beta}}^{\widehat{\beta}} \left[\left(\mu + \sqrt{-\Phi_M(\widehat{\beta}, s)} \right)^2 + \frac{\alpha - s}{\tau^*(s) + \phi} + \frac{\beta + s}{\phi} + c\tau^*(s) \right] f_\beta(s) ds.$$

Thus his total payoff including stage 1 is:

$$V(\beta, \widehat{\beta}) = \Pi(\beta, \widehat{\beta}) - \lambda \left[\frac{\alpha + \beta}{\tau^S(\widehat{\beta}) + \phi} + c\tau^S(\widehat{\beta}) \right],$$

where λ reflects the relative importance of the payoffs in stages 1 and 3.

To determine the candidate's optimal choice of $\widehat{\beta}$ given the signaling schedule τ^S , we take the derivative of V with respect to $\widehat{\beta}$ and set it equal to zero: $\frac{\partial}{\partial \widehat{\beta}} V(\beta, \widehat{\beta}) = 0$. In a fully separating equilibrium the solution is $\widehat{\beta} = \beta$, which implies $\frac{\partial}{\partial \widehat{\beta}} V(\beta, \widehat{\beta}) \Big|_{\widehat{\beta}=\beta} = 0$. Notice that we can rewrite this equilibrium condition as follows:

$$(25) \quad \frac{\partial \Pi(\beta, \widehat{\beta})}{\partial \widehat{\beta}} \Big|_{\widehat{\beta}=\beta} = \lambda c \left[1 - \left(\frac{\tau^*(\beta) + \phi}{\tau^S(\beta) + \phi} \right)^2 \right] \frac{\partial \tau^S}{\partial \widehat{\beta}}.$$

Now observe that equation 25 is a nonlinear first-order differential equation. As usual, the equilibrium leaves the choice of the "worst" type undistorted, so we also have a boundary condition, $\tau^S(\overline{\beta}) = \tau^*(\overline{\beta})$.

A somewhat unconventional feature of this differential equation is that the coefficient of $\frac{\partial \tau^S}{\partial \widehat{\beta}}$ is zero at the initial condition, which renders $\frac{\partial \tau^S}{\partial \widehat{\beta}} \Big|_{\widehat{\beta}=\beta}$ undefined. We can finesse this difficulty by reversing the mathematical roles of τ^S and β , treating equation 25 as a differential equation for the function $\beta^S(\tau)$, with initial condition $\beta^S(\tau) = \overline{\beta}$ for $\tau = \tau^*(\overline{\beta})$. In that case, we have $\frac{\partial \beta^S}{\partial \tau} \Big|_{\tau=\tau^*(\overline{\beta})} = 0$. Our next result describes some properties of the solution.

Proposition 9. $\beta^S(\tau)$ is strictly increasing in τ with $\tau < \tau^*(\beta^S(\tau))$ for $\tau < \tau^*(\bar{\beta})$, and strictly decreasing in τ with $\tau > \tau^*(\beta^S(\tau))$ for $\tau > \tau^*(\bar{\beta})$.

This proposition offers us two candidates for the separating function τ^S : we can invert the solution β^S restricting attention either to $\tau < \tau^*(\bar{\beta})$, which yields an upward-sloping function $\tau^1(\beta)$ satisfying $\tau^1(\beta) < \tau^*(\beta)$, or to $\tau > \tau^*(\bar{\beta})$, which yields a downward-sloping function $\tau^2(\beta)$. Given our assumptions, there is no guarantee that either solution is globally incentive-compatible. Our next result tells us that only τ^1 can serve as the separating function τ^S . We show that a sufficiently large value of λ – which one can interpret (for example) as a limit on the likelihood with which lower officeholders expects to run for higher office – guarantees global incentive compatibility of τ^1 . The proof also supplies a relatively simple analytic condition that one can check after solving any parametrized version of the model numerically; see the Appendix. In contrast, τ^2 is never globally incentive compatible. Accordingly, we will henceforth associate τ^S with τ^1 .

Proposition 10. *There exists λ^* such that for $\lambda > \lambda^*$, τ^1 is globally incentive-compatible. τ^2 is never globally incentive compatible.*

Together, propositions 9 and 10 imply that $\tau^S(\beta) < \tau^*(\beta)$. Thus, politicians signal by acting more decisive than they actually are. In addition, we learn that τ^S must be strictly increasing in β , which raises the possibility that the non-negativity constraint ($\tau > 0$) may bind. We will eliminate this possibility by assuming, where necessary, that ϕ is sufficiently small.

Figure 2 depicts a typical equilibrium. We produced the figure by numerically solving a parametrized version of our model. The horizontal axis describes the politician’s type, and the vertical axis describes the time take to make a decision. For this case, we have also verified global incentive-compatibility.³⁷ The figure shows three functions relating the decision time, τ , to the politician’s knowledge of his idiosyncratic preferences, β : the ideal from the politician’s perspective (τ^*), the ideal from a voter’s perspective (τ^v), and the signaling equilibrium (τ^S). Because we have assumed the cost of delay, c , is the same for all voters, the ideal τ^v does not depend on the voter’s identity.

For the moment, focus on the curves representing the voters’ and politician’s ideals. One important feature of the figure is that the former lies below the latter. In other words, absent heterogeneity in c , voters regard all non-strategic politicians as hesitant. The intuition is simply that voters bear additional risk when politicians take extra time to resolve

³⁷The figure was generated using $\mu = 0$, $\phi = 0.001$, $\alpha = 1$, $c = 25$, $\beta = 0$, and $\bar{\beta} = 0.2$. β_P is distributed as per a truncated normal distribution with a mean of 0.2 and a standard deviation of 0.0667.

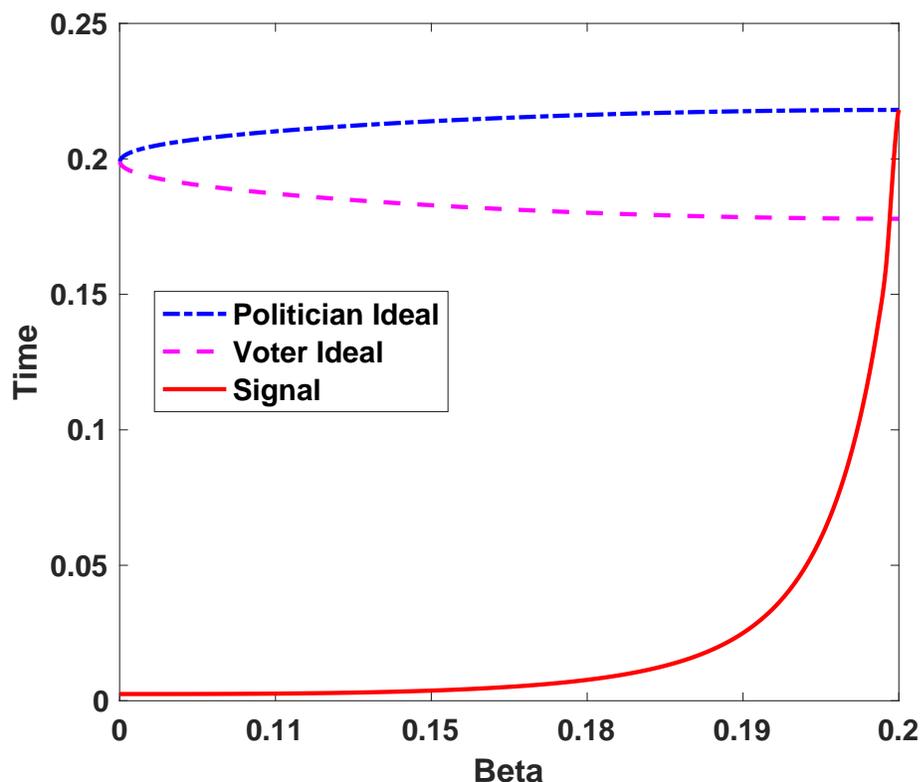


FIGURE 2. Equilibrium Versus Voter and Politician Optima With Signaling of β

the fuzziness of their views. Notice also that the curve representing the politician's ideal slopes upward, while the one representing the voter's ideal slopes downward. Intuitively, if the politician starts out with less precise knowledge of his own preferences, incremental time spent deliberating will lead to a greater reduction in the risk he bears, but a greater increase in the risk born by the voter; hence the ideal duration of the deliberation period increases with β for the politician and decreases for the voter.

As in Section 5, signaling does not necessarily improve voter welfare. Signaling is a rather poor solution for politicians' tendency to delay because the signaling incentive is smallest where the need for a corrective influence is greatest, and greatest where that need is smallest. That pattern is evident from the figure. Signaling has no effect on politicians with the largest values of β , for whom the gap between the voter and politician ideal is greatest. Moreover, its cumulative effect on politicians with the lowest values of β , for whom that gap is smallest, can be enormous, causing them to spend little or no time

pondering the common good even when they should. In the figure, the signaling curve is so steeply sloped that it crosses the voter ideal curve. Politicians to the right of the crossing remain hesitant from the voter's perspective, but those to the left become hasty, overshooting the voter's ideal. Thus the overall impact on voter welfare can be positive or negative, depending on the size of the signaling effect and the distribution of politician types – specifically, whether it is skewed toward those with relatively good knowledge of idiosyncratic preferences who overcorrect, or those with relatively poor knowledge of idiosyncratic preferences who undercorrect.

Because we have not endogenized the decision to become a politician, we cannot draw formal conclusions about the distribution of candidate types. However, we can still provide some insight concerning this issue. Our next result establishes that higher quality politicians receive higher expected payoffs in equilibrium.

Proposition 11. $V(\beta, \beta)$ is decreasing in β .

Were we to endogenize the distribution of candidate types, for example by appending the familiar citizen-candidate apparatus to our model, proposition 11 suggests that the distribution of politicians would likely be skewed toward types with low values of β that the voters' prefer. However, coupled with the preceding analysis, this additional observation raises the possibility that the political system may encourage most politicians to act with excessive haste in order to project decisiveness.

APPENDIX D. ONLINE APPENDIX: PROOFS FROM APPENDIX B

Our proofs for the equilibrium for the game wherein agents signal knowledge of their idiosyncratic preferences are similar to the arguments made when analyzing the model wherein agents signal an aversion to delay. For our next result, we rewrite the first-order condition, equation (10), so that it describes a differential equation for β in terms of τ , rather than the other way around:

$$(26) \quad \frac{\partial \beta^S}{\partial \tau} = \lambda c \left[\frac{\partial \Pi(\beta^S(\tau), \widehat{\beta})}{\partial \widehat{\beta}} \Big|_{\widehat{\beta}=\beta^S(\tau)} \right]^{-1} \left[1 - \left(\frac{\tau^*(\beta^S(\tau)) + \phi}{\tau + \phi} \right)^2 \right].$$

The following lemma ensures that the right hand side of equation 26 is well-defined:

Lemma 6. *There exist finite $d_L, d_U > 0$ such that $d_L < -\frac{\partial \Pi(\beta, \widehat{\beta})}{\partial \widehat{\beta}} \Big|_{\widehat{\beta}=\beta} < d_U$ for all $\beta \in [\underline{\beta}, \overline{\beta}]$.*

Proof. Taking the derivative, we have

$$(27) \quad -\frac{\partial \Pi(\beta, \widehat{\beta})}{\partial \widehat{\beta}} \Big|_{\widehat{\beta}=\beta} = 2\beta \left[\frac{1}{\phi} - \frac{1}{\tau^*(\beta) + \phi} \right] f_\beta(\beta) \\ - \int_{\underline{\beta}}^{\overline{\beta}} \left(\frac{\mu}{\sqrt{\Phi_M(\beta, s)}} - 1 \right) \frac{\partial \Phi_M(\beta, s)}{\partial \widehat{\beta}_1} f_\beta(s) ds \\ - \int_{\underline{\beta}}^{\overline{\beta}} \left(\frac{\mu}{\sqrt{-\Phi_M(\beta, s)}} + 1 \right) \frac{\partial \Phi_M(\beta, s)}{\partial \widehat{\beta}_1} f_\beta(s) ds.$$

Let

$$d_L \equiv 2\underline{\beta} \left[\frac{1}{\phi} - \frac{1}{\tau^*(\underline{\beta}) + \phi} \right] \underline{f}_\beta > 0,$$

where \underline{f}_β is the lower bound on density. By construction, the first term in equation 27 exceeds d_L . The second and third terms are positive because (i) $\sqrt{|\Phi_M(\beta, s)|} < \mu$ by assumption, and (ii) Φ_M is decreasing in its first argument. Thus, the sum exceeds d_L .

Next, let

$$d_U^1 \equiv 2\overline{\beta} \left[\frac{1}{\phi} - \frac{1}{\tau^*(\overline{\beta}) + \phi} \right] \overline{f}_\beta > 0,$$

where \overline{f}_β is the upper bound on density. By construction, d_U^1 exceeds the first term in equation 27. From the proof of Lemma 2, it is easily seen that there exists finite values $\underline{e}, \overline{e} > 0$ such that $\underline{e} < -\frac{\partial \Phi_M(\underline{\beta}_1, \underline{\beta}_2)}{\partial \widehat{\beta}_1}, \frac{\partial \Phi_M(\underline{\beta}_1, \underline{\beta}_2)}{\partial \widehat{\beta}_2} < \overline{e}$ for all $\widehat{\beta}_1, \widehat{\beta}_2 \in [\underline{\beta}, \overline{\beta}]$. Accordingly,

$$-\int_{\underline{\beta}}^{\overline{\beta}} \left(\frac{\mu}{\sqrt{\Phi_M(\beta, s)}} - 1 \right) \frac{\partial \Phi_M(\beta, s)}{\partial \widehat{\beta}_1} f_\beta(s) ds \leq -\frac{\overline{e}\overline{f}}{\underline{e}} \int_{\underline{\beta}}^{\overline{\beta}} \left(\frac{\mu}{\sqrt{\Phi_M(\beta, s)}} - 1 \right) \frac{\partial \Phi_M(\beta, s)}{\partial s} ds \\ = -\frac{\overline{e}\overline{f}}{\underline{e}} \left[2\mu\sqrt{\Phi_M(\beta, s)} - \Phi_M(\beta, s) \right]_{\underline{\beta}}^{\overline{\beta}} \\ \leq \frac{\overline{e}\overline{f}}{\underline{e}} \left[2\mu\sqrt{\Phi_M(\underline{\beta}, \overline{\beta})} + \Phi_M(\underline{\beta}, \overline{\beta}) \right] \\ \equiv d_U^2,$$

where we have used the bounds on the derivatives of Φ_M to switch the partial derivative in the first line. A completely analogous argument yields an upper bound, d_H^3 , on the final term in equation 27. To complete the proof, we take $d_U = d_U^1 + d_U^2 + d_U^3$. \square

Proposition 9. $\beta^S(\tau)$ is strictly increasing in τ with $\tau < \tau^*(\beta^S(\tau))$ for $\tau < \tau^*(\bar{\beta})$, and strictly decreasing in τ with $\tau > \tau^*(\beta^S(\tau))$ for $\tau > \tau^*(\bar{\beta})$.

Proof. Substituting the initial condition into equation 26, we obtain $\left. \frac{\partial \beta^S}{\partial \tau} \right|_{\tau=\tau^*(\bar{\beta})} = 0$, as claimed in the text. Using equation 26 along with Lemma 6, we see that $\frac{\partial \beta^S}{\partial \tau} < 0$ when $\tau > \tau^*(\beta^S(\tau))$, and $\frac{\partial \beta^S}{\partial \tau} > 0$ when $\tau < \tau^*(\beta^S(\tau))$.

First consider $\tau > \tau^*(\bar{\beta})$. Starting at $\tau^*(\bar{\beta})$, a small increase in τ has a negligible effect on β^S , leaving us with $\tau > \tau^*(\beta^S(\tau))$, and hence $\frac{\partial \beta^S}{\partial \tau} < 0$. As τ increases, the slope remains strictly negative unless we reach a point at which $\tau = \tau^*(\beta^S(\tau))$. But that is impossible, because $\tau^*(\beta) < \tau^*(\bar{\beta})$ for all $\beta < \bar{\beta}$.

Next consider $\tau < \tau^*(\bar{\beta})$. Starting at $\tau^*(\bar{\beta})$, a small decrease in τ has a negligible effect on β^S , leaving us with $\tau < \tau^*(\beta^S(\tau))$, and hence $\frac{\partial \beta^S}{\partial \tau} > 0$. As τ increases, the slope remains strictly positive unless we reach a value τ' at which $\tau' = \tau^*(\beta^S(\tau'))$. But that is impossible, because (i) $\frac{\partial \tau^*}{\partial \beta}$ is strictly positive and bounded away from 0 on $[\underline{\beta}, \bar{\beta}]$, and (ii) from (26) and lemma 6, $\frac{\partial \beta^S}{\partial \tau}$ would converge to zero as $(\tau, \beta^S(\tau))$ converged to $(\tau', \tau^*(\beta^S(\tau')))$; thus, further reductions in τ would widen the gap between $\beta^S(\tau)$ and $(\tau^*)^{-1}(\tau)$. \square

Proposition 10. *There exists λ^* such that for $\lambda > \lambda^*$, τ^1 is globally incentive-compatible. τ^2 is never globally incentive compatible.*

Proof. We begin by showing that τ^1 converges to τ^* , and establishing a lower bound on $\frac{\partial \tau^1}{\partial \beta}$. To this end, we define $R(\lambda) = \max_{\beta \in [\underline{\beta}, \bar{\beta}]} \frac{\tau^*(\beta) + \phi}{\tau^1(\beta) + \phi} > 1$. Continuity of τ^* and τ^1 ensures existence of the maximum.

Lemma 7. *For any λ and all $\beta \in [\underline{\beta}, \bar{\beta}]$, we have $\frac{\partial \tau^1}{\partial \beta} > \frac{d_L}{\lambda c(R(\lambda)^2 - 1)}$. Furthermore, $\lim_{\lambda \rightarrow \infty} R(\lambda) = 1$.*

Proof. Fix a value of λ . From equation 25 and lemma 6, we have $-d_L \leq \lambda c(1 - R(\lambda)^2) \frac{\partial \tau^1}{\partial \beta}$, or equivalently $\frac{\partial \tau^1}{\partial \beta} > \frac{d_L}{\lambda c(R(\lambda)^2 - 1)}$, for all $\beta \in [\underline{\beta}, \bar{\beta}]$, which establishes the first part of the lemma.

For the second part of the lemma, we begin by observing that $\frac{\partial \tau^*}{\partial \beta} = \frac{1}{2\sqrt{c(\alpha + \beta)}}$, which is bounded below by $g \equiv \frac{1}{2\sqrt{c(\alpha + \bar{\beta})}}$.

Now fix $r > 1$, and define $\lambda_r \equiv r \frac{d_H}{g(r^2 - 1)}$. From equation 25 and the definitions of d_H and g , it follows that, if $\lambda > \lambda_r$ and $\frac{\tau^*(\beta) + \phi}{\tau^1(\beta) + \phi} \geq r$, we would have $r \frac{\partial \tau^1}{\partial \beta} < \frac{\partial \tau^*}{\partial \beta}$.

Next we show that, for all $\lambda > \lambda_r$, we have $R(\lambda) < r$. Suppose on the contrary that, for such λ , there is some β for which $\frac{\tau^*(\beta) + \phi}{\tau^1(\beta) + \phi} \geq r$. Let β' be the largest value for which

$\frac{\tau^*(\beta') + \phi}{\tau^1(\beta') + \phi} = r$. Differentiating, we obtain

$$\begin{aligned} \left. \frac{d}{d\beta} \left(\frac{\tau^*(\beta) + \phi}{\tau^1(\beta) + \phi} \right) \right|_{\beta=\beta'} &= \frac{\frac{d\tau^*}{d\beta} (\tau^1(\beta') + \phi) - \frac{d\tau^1}{d\beta} (\tau^*(\beta') + \phi)}{(\tau^1(\beta') + \phi)^2} \\ &= \frac{\frac{d\tau^*}{d\beta} - \frac{d\tau^1}{d\beta} r}{\tau^1(\beta') + \phi} \\ &> 0. \end{aligned}$$

It follows that a small increase in β from β' would result in $\frac{\tau^*(\beta) + \phi}{\tau^1(\beta) + \phi} > r$. But we know that $\frac{\tau^*(\bar{\beta}) + \phi}{\tau^1(\bar{\beta}) + \phi} = 1 < r$. Consequently, there would have to be some $\beta'' \in (\beta', \bar{\beta})$ for which $\frac{\tau^*(\beta'') + \phi}{\tau^1(\beta'') + \phi} = r$. But that contradicts the definition of β' .

We conclude that, for any $r > 1$, we have $R(\lambda) < r$ for all $\lambda > \lambda_r$. The lemma follows directly. \square

Lemma 8. *There exists λ^* such that for all $\lambda > \lambda^*$, we have $\frac{\partial^2 V(\beta, \hat{\beta})}{\partial \beta \partial \hat{\beta}} > 0$ for all $\beta \in [\underline{\beta}, \bar{\beta}]$.*

Proof. Using equation 3 to substitute for $\tau^*(\beta)$ in equation 24 and differentiating, we obtain

$$\begin{aligned} (28) \quad \frac{\partial^2 V(\beta, \hat{\beta})}{\partial \hat{\beta} \partial \beta} &= \lambda \left(\frac{1}{\tau^S(\hat{\beta}) + \phi} \right)^2 \frac{\partial \tau^1}{\partial \hat{\beta}} - \left(\frac{1}{\phi} - \sqrt{\frac{c}{\alpha + \beta}} \right) f_\beta(\hat{\beta}) \\ &> \left(\frac{1}{\tau^*(\bar{\beta}) + \phi} \right)^2 \frac{d_L}{c(R(\lambda)^2 - 1)} - \left(\frac{1}{\phi} - \sqrt{\frac{c}{\alpha + \bar{\beta}}} \right) \bar{f}_\beta, \end{aligned}$$

where \bar{f}_β is an upper bound on the density of β , and the inequality makes use of Lemma 4. Observe that (i) $\left(\frac{1}{\tau^*(\bar{\beta}) + \phi} \right)^2$ is strictly positive and independent of λ , (ii) $\frac{d_L}{c(R(\lambda)^2 - 1)}$ is strictly positive and, by Lemma 4, converges to $+\infty$ as $\lambda \rightarrow +\infty$, and (iii) $\left(\frac{1}{\phi} - \sqrt{\frac{c}{\alpha + \bar{\beta}}} \right) \bar{f}_\beta$ is finite. Accordingly, the entire expression is strictly positive for λ sufficiently large. \square

Now we prove the first part of the proposition. In light of the fact that $\frac{\partial^2 V(\beta, \hat{\beta})}{\partial \beta \partial \hat{\beta}} > 0$, the first-order condition for claiming to be of type $\hat{\beta}$, $\frac{\partial V(\beta, \hat{\beta})}{\partial \hat{\beta}} = 0$, can be satisfied by at most one type, which is by construction $\beta = \hat{\beta}$. Consequently the optimal value of $\hat{\beta}$ for each β must be β , $\underline{\beta}$, or $\bar{\beta}$. We can rule out $\underline{\beta}$ on the grounds that $\left. \frac{\partial V(\beta, \hat{\beta})}{\partial \hat{\beta}} \right|_{\beta, \hat{\beta}=\underline{\beta}} = 0$ and $\frac{\partial^2 V(\beta, \hat{\beta})}{\partial \hat{\beta} \partial \beta} > 0$; similarly for $\bar{\beta}$. We conclude that τ^S is globally incentive-compatible.

Turning to the second part of the proposition, we know that $\lim_{\beta \rightarrow \bar{\beta}} \frac{\partial \tau^2}{\partial \beta} = -\infty$. Accordingly, the analog of equation 28 tells us that $\frac{\partial^2 V(\beta, \hat{\beta})}{\partial \hat{\beta} \partial \beta} < 0$ for β and $\hat{\beta}$ in a neighborhood

of $\bar{\beta}$. But in that case, local incentive compatibility is violated: for β_1 and β_2 close to $\bar{\beta}$, the fact that $V(\beta_1, \beta_1) > V(\beta_1, \beta_2)$ would imply $V(\beta_2, \beta_1) > V(\beta_2, \beta_2)$. \square

Remark 1. *The proof also supplies a sufficient condition for global compatibility that we can check after solving for τ^S numerically: for all $\beta, \hat{\beta} \in [\underline{\beta}, \bar{\beta}]$:*

$$\Omega(\beta, \hat{\beta}) \equiv \frac{1}{\left(\tau^S(\hat{\beta}) + \phi\right)^2} \frac{d\tau^S(\hat{\beta})}{d\hat{\beta}} - \left(\frac{1}{\phi} - \frac{1}{\tau^*(\beta) + \phi}\right) f_{\beta}(\hat{\beta}) > 0.$$

Proposition 11. *$V(\beta, \beta)$ is decreasing in β .*

Proof. Consider $\beta < \beta'$ and the following decomposition of $V(\beta, \beta) - V(\beta', \beta')$

$$V(\beta, \beta) - V(\beta', \beta') = [V(\beta, \beta) - V(\beta, \beta')] + [V(\beta, \beta') - V(\beta', \beta')].$$

The first term is positive since incentive compatibility requires $V(\beta, \beta) \geq V(\beta, \beta')$. Using some algebra we find

(29)

$$\begin{aligned} V(\beta, \beta') - V(\beta', \beta') &= - \int_{\beta'}^{\bar{\beta}} \left[\frac{\alpha + \beta}{\tau^*(\beta) + \phi} - \frac{\alpha + \beta'}{\tau^*(\beta') + \phi} + c[\tau^*(\beta) - \tau^*(\beta')] \right] f_{\beta}(s) ds \\ &\quad - \int_{\underline{\beta}}^{\beta'} \left[\frac{\beta - \beta'}{\phi} \right] f_{\beta}(s) ds - \lambda \frac{\beta - \beta'}{\tau^S(\beta') + \phi}. \end{aligned}$$

Using the fact that

$$\frac{\alpha + \beta}{\tau^*(\beta) + \phi} = \sqrt{c(\alpha + \beta)} = c \sqrt{\frac{\alpha + \beta}{c}} = c(\tau^*(\beta) + \phi),$$

we obtain the following

$$\begin{aligned} (30) \quad V(\beta, \beta') - V(\beta', \beta') &= -2c \int_{\beta}^{\bar{\beta}} [\tau^*(\beta) - \tau^*(\beta')] f_{\beta}(s) ds \\ &\quad - \int_{\underline{\beta}}^{\beta} \left[\frac{\beta - \beta'}{\phi} \right] f_{\beta}(s) ds - \lambda \frac{\beta - \beta'}{\tau^S(\beta') + \phi}. \end{aligned}$$

Since $\beta < \beta'$ implies $\tau^*(\beta) < \tau^*(\beta')$, all of these terms are positive, so we have $V(\beta, \beta') - V(\beta', \beta') \geq 0$. Therefore $V(\beta, \beta) \geq V(\beta', \beta')$. \square