Motivated reasoning and democratic accountability*

Andrew T. Little† Keith E. Schnakenberg‡ Ian R. Turner§

June 2020

Abstract

Standard political agency models assume voters form accurate beliefs about how politicians are performing. However, substantial evidence from political behavior research indicates that voters have "directional motives" beyond accuracy. These results are often taken as evidence that voters will not be able to hold politicians accountable for their actions. We probe this conclusion by formalizing a model of accountability where voters form beliefs with both directional and accuracy motives. We classify the effects of directional motives into (1) *divergence*, which causes people with different preferences to hold different beliefs about relevant facts, and (2) *desensitization*, which leads to a weaker relationship between incumbent performance and voter beliefs. We show that divergence has an ambiguous impact on politician behavior but desensitization uniformly decreases incentives to perform well in office. A key implication of our analysis is that motivated reasoning generally does harm democratic accountability, but this is not because partisans get more polarized, but because all voters become less sensitive to changes in incumbent performance. We also explore the implications for empirical work which examines the relationship between performance indicators and incumbent vote shares. While directional motivated reasoning always weakens the strength of this relationship, we can't necessarily infer that accountability is diminished.

Preliminary draft.
Comments appreciated.

*Many thanks to Bethany Albertson, Taylor Carlson, Greg Huber, Chris Lucas, William Nomikos, Dave Siegel, Carly Wayne, Stephane Wolton, and audience members at APSA 2019, SPSA 2020, Yale, LSE, and Texas A&M for helpful feedback.

†Assistant Professor of Political Science, UC Berkeley. Contact: andrew.little@berkeley.edu

‡Assistant Professor of Political Science, Washington University in St. Louis. Contact: keschnak@wustl.edu.

§Assistant Professor of Political Science, Yale University. Contact: ian.turner@yale.edu.

A rich and accurate model of voter behavior needs to account for sensible voter responsiveness but also these [psychological] biases, identifying the conditions under which retrospective voting achieves effective democratic accountability and when it fails to do so.

Healy and Malhotra (2013)

Voters are often misinformed about political facts. Furthermore, biases in voter beliefs are systematic in a way that suggests they may interpret information using directional motivated reasoning that reinforces their preferences for the candidates or policies they already preferred (Taber and Lodge 2006, Redlawsk 2002). These findings lead some to worry that standard models of electoral accountability do not describe real electorates and that voters are in fact not competent enough to induce desirable behavior by politicians (e.g., Achen and Bartels 2017). However, the implications of voters' motivated reasoning for democratic performance are not well understood since improving voters' information about politicians' performance need not always improve democratic performance (Ashworth and Bueno do Mesquita 2014). The voter-centric orientation of the political psychology literature and the focus of political economy models on rational voters leaves us without answers to some basic questions. Is directional motivated reasoning a concern for democracy? If so, which motives and patterns of beliefs are of greatest concern? What is the connection, if any, between the effects of motivated reasoning on the accountability of incumbent politicians and electoral outcomes?

To answer these questions, we analyze a variant of a standard ("career concerns") political economy model of an election. An incumbent politician decides how hard to work on behalf of voters. The incumbent's "performance" in office is increasing in this effort, as well as her competence and a random shock. Voters observe the performance, form a conclusion about the incumbent's competence, and prefer to re-elect more competent representatives.[1] As long as these conclusions are increasing in incumbent performance, she will have an incentive to put in effort to increase her chance of re-election. The accountability mechanism works better when these incentives for effort are strong.

[1] We use the term *conclusion* to refer to an estimate of incumbent competence. We say conclusion rather than belief to distinguish this estimate from a full probability distribution over types.

In the standard model, voters form their beliefs about the incumbent's performance according to Bayes's rule. Or, in the terminology from the behavioral literature we draw on, voters only have *accuracy motives*. Our key innovation is to assume that in addition to accuracy motives, voters also have *directional motives*: conclusions about the incumbent politician that they like more than others, independent of performance. Our model highlights two potential behavioral effects of this kind of motivated reasoning and spells out the implications for both politicians' performance in office and aggregate incumbent vote share.[2]

Divergence and desensitization. The first effect of motivated reasoning is *divergence* of the voters' conclusions about the incumbent. Voters with different preexisting affinity towards or against the incumbent form different conclusions about her competence, and increasing divergence refers to the phenomenon where these differences are increasing in voters' directional motives. Divergence is a theoretical analog to many empirical results which have long been seen as a serious risk to democratic accountability. For instance, members of different parties have different perceptions of economic indicators (Bartels 2002), of whether weapons of mass destruction were found in Iraq (Jacobson 2010), or of what proportion of the federal budget is allocated to welfare programs (Kuklinski et al. 2000).

Our model also formalizes an important caveat to these findings raised by some scholars of partisanship: divergence does not imply that voters are unresponsive to information, and in fact they may all respond to changes in incumbent performance in parallel (Gerber and Green 1999, Green, Palmquist and Schickler 2004). That is, in our conception of motivated reasoning, no matter how much voter conclusions diverge for a fixed performance level, all individual conclusions respond to performance in the 'correct' direction – positive (negative) performance weakly improves (harms) conclusions of incumbent competence. That is, divergence is *not* driven by a "backlash" effect where, say, those with an affinity towards the incumbent respond to negative information about her performance by doubling down on their support (consistent with what empirical research with reliable research designs and samples sizes has found, see Guess and Coppock 2018).

[2]We also analyze how motivated reasoning affects whether less competent politicians are chosen when voters have stronger directional motives in section 3.1. The analysis of selection is more complicated, but broadly similar.

However, this does not imply that motivated reasoning has no impact on how voters respond to information. The second effect of motivated reasoning we identify, *desensitization*, is a weakening of this response. Independent of divergence effects, stronger directional motives can weaken the connection between actual performance and posterior estimates of incumbent quality. Desensitization captures, for example, the description of "cautious Bayesians" – voters whose beliefs move toward the correct beliefs but not far enough – in Hill (2017). This effect is consistent with empirical patterns that suggest information about the economy impacting vote choice has weakened in recent years (e.g., Donovan et al. 2019, Freeder 2019). Though desensitization may be caused by the same general processes as divergence, for instance if two sides are locked into opposing beliefs and are unwilling to change, the two processes need not go hand in hand. In fact, our first set of results show how several natural formulations of voters' accuracy and directional motives lead to different combinations of belief divergence and belief desensitization.

Motivated reasoning and accountability. Our core theoretical contention is that desensitization effects have more uniform consequences for democratic performance than do divergence effects. The reasoning for this is two-fold. First, the mechanism through which voters beliefs affect politicians' incentives for effort is by changing the relationship between performance in office and the likelihood of re-election. Desensitization leads voters to have a weaker response to changes in performance, which therefore causes lower effort by politicians. Divergence alone need not have this effect: if some voters evaluate the incumbent more favorably than others but each respond to changes in performance in the same way, politicians' behavior may be approximately the same as with fully Bayesian voters. Second, electoral incentives depend mostly on the behavior of voters in the middle. If centrist voters are not inclined toward one side or the other, divergence will not have a substantial effect on their conclusions and, therefore, will not affect politicians' behavior. However, desensitization affects all voters, including centrists. When all voters become less responsive to incumbent performance, this can severely reduce politicians' incentives to work hard on their behalf.

This observation leads to the following contention: while motivated reasoning may undermine

politicians' incentives to work on behalf of voters, the reason why this is true is not what one might expect. Our results suggest that the key problem is not committed partisans holding systematically divergent beliefs. Instead, the threat to democratic accountability from motivated reasoning is that even moderate voters become less responsive to new information.

Motivated reasoning and electoral outcomes. We also tie our results to the larger empirical literature which studies the relationship between performance indicators (economic growth, crime, educational outcomes) and incumbent vote shares. This literature often interprets the strength of this relationship as a proxy for the strength of politician incentives to provide good outcomes. Further, some recent research argues that a decline in this relationship in the U.S. is driven by increased partisan motivated reasoning (e.g., Freeder 2019). Our model highlights that the degree to which this kind of inference is warranted depends on the kind of motivated reasoning in question. We show that *both* divergence and desensitization will weaken the correlation between performance and vote shares, as this relationship depends on more than just the median voter. However, recall that only desensitization unambiguously decreases incumbent performance. Thus, *if the primary effect of motivated reasoning is that it drives belief divergence then it can weaken the relationship between performance and vote shares while having a minimal effect on electoral accountability.*

Summary. We view our paper as making three primary contributions. First, we make a methodological contribution by integrating a game-theoretic model with a model of motivated reasoning (section 2). The model of belief formation is based on Little (2019), which conceptualizes motivated reasoning as an optimization problem balancing accuracy motivations and directional motivations (Kunda 1990).[3] Ours is the first paper to incorporate this model of beliefs into a game-theoretic or decision-theoretic model, which we see as a promising avenue for synthesizing work in political behavior and political economy. A concrete aspect of this contribution is to decompose the effects of motivated reasoning into belief divergence and belief desensitization, which could apply well beyond the particular accountability model we study.

[3]Thaler (2019) presents a related model of motivated reasoning in which agents update beliefs using Bayes's rule but treat their directional motives as an extra signal, and conducts experiments where this kind of motivated reasoning can be disentangled from standard Bayesian explanations.

Second, within the context of democratic accountability, our model helps to clarify normative implications of contemporary work in political psychology and political behavior. Our model not only explains when voter biases might matter for electoral accountability, but also helps explain *which* biases affect accountability and suggests a way to categorize different findings in the literature to help think through their potential electoral effects (section 3). Moreover, this decomposition of effects also illustrates how motivated reasoning can affect aggregate electoral outcomes without negatively affecting the accountability incentives provided by elections (section 4).

Third, we show that the effects of motivated reasoning on voters and politicians parallel effects which are studied in "standard" models (section 5). In other words, while motivated reasoning matters, these effects are not qualitatively different than the effects of more standard parameters like heterogeneity of voter preferences, the quality of the information environment, and how much voters weigh performance versus other non-performance-related factors. The risks to effective democratic governance driven by motivated reasoning are important to consider, but they are not uniquely troubling in terms of voters' ability to hold their representatives accountable.

1 Related literature

To develop a model of democratic accountability that explicitly incorporates motivated reasoning, we build on a voluminous theoretical literature in political economy that studies electoral accountability (e.g., Fearon 1999).[4] This literature has provided insight into how and when elections can produce high quality governance. The ability of elections to produce good government follows from citizens' ability to select quality politicians and hold incumbents accountable through the threat of removing them from office through voting. Typically these two effects depend on voters updating their beliefs about incumbent quality based on performance information. This improves selection by increasing the likelihood that good politicians are reelected and improves performance incentives because incumbents want to produce signals that they are high quality.

[4]Among many others, see also Ashworth (2005), Ashworth and Bueno de Mesquita (2008), Canes-Wrone, Herron and Shotts (2001), Fox (2007), Fox and Jordan (2011), Gordon, Huber and Landa (2007), Li, Sasso and Turner (2020), Maskin and Tirole (2004), Schnakenberg (2020). The specific model we adapt is closest to Persson and Tabellini (2002). Ashworth (2012) and Duggan and Martinelli (2017) provide reviews of the literature.

Empirical research on electoral accountability is often more skeptical about the ability of citizens to effectively hold their representatives to account (e.g., Lupia and McCubbins 1998). Much of this work has focused on whether voters are sufficiently informed (Delli Carpini and Keeter 1996, Popkin 1991) or sufficiently rational (Achen and Bartels 2017, Healy, Malhotra and Mo 2010, Woon 2012) to hold politicians accountable in ways predicted by the models noted above.[5] Most closely related to our aims are studies arguing that directional motivated reasoning leads voters to form biased beliefs about politics (e.g., Lodge and Taber 2013, Taber and Lodge 2006),[6] often caused by partisan attachments coloring voter perceptions (Bolsen, Druckman and Cook 2014).[7] These deviations may be particularly harmful to accountability if voters with different partisanship or ethnicity only react to information they are already predisposed to believe (Adida et al. 2017). Redlawsk (2002) characterizes motivated reasoning specifically as "a direct challenge to the notion of candidate evaluation as a Bayesian updating process in which voters readily modify their prior expectations based on the value of new information" (pg. 1041).

Yet, an established result in political science is that incumbent politicians tend to fare better electorally when government or the economy is generally performing well (e.g., De Benedictis-Kessner and Warshaw N.d., Fair 1978, Ferraz and Finan 2008, Kramer 1971, Lewis-Beck and Stegmaier 2000).[8] This seems to be consistent with voters' ability to hold their representatives to account, largely in line with the 'standard' political economy model predictions.

A broad reading of these literatures suggests therefore that voters are individually flawed but aggregate outcomes reward good performance in predictable ways. Ashworth and Fowler (2019) provide one way to reconcile these findings, arguing that the presence of flawed *individual* voters

[5] See Anderson (2007) and Healy and Malhotra (2013) for reviews.

[6] Among many others, see also Bisgaard (2019), Kahan (2012), Lau and Redlawsk (2006), Nyhan and Reifler (2010), Prior et al. (2015), Redlawsk, Civettini and Emmerson (2010), Strickland, Taber and Lodge (2011), Taber, Cann and Kucsova (2009). Druckman and McGrath (2019) provides a review in the context of beliefs about climate change.

[7] Though see Bullock and Lenz (2019) which suggests survey answers may sometimes reflect "cheerleading" rather than sincere and strongly held beliefs. For example, Bullock et al. (2015) provide evidence that these differences become smaller when respondents are paid to give accurate answers. De Vries, Hobolt and Tilley (2018) also show that presentation of unambiguous real-world conditions increase accurate answers across partisans. Fowler (2020) and Orr and Huber (2019) provide evidence that policy views underlie partisan disagreement more than is often appreciated.

[8] For research with similar results see also Alt, Bueno de Mesquita and Rose (2011), Eggers et al. (2014), Fowler et al. (2016), Hall (2015), Hirano et al. (2015), Lenz (2013), Payson (2017).

does not preclude *aggregate electorates* from enforcing democratic accountability. So long as the effective representative voter in the public is able to do so all other voters can suffer from behavioral biases that would appear to undermine electoral accountability. We start with a premise even closer to empirical critique of accountability models— that no voters are fully rational—and reach a similar conclusion about the aggregate consequences of these biases. By explicitly incorporating motivated reasoning processes into the model, however, we also provide more detailed results on the consequences of different voter biases.

Methodologically, we contribute to a burgeoning literature in political economy that incorporates psychological concepts into formal models.[9] For example, Ashworth and Bueno do Mesquita (2014) show how voters who fail to form correct conjectures about politician effort affect the equilibrium behavior and welfare properties of a similar model. Diermeier and Li (2017) study an accountability model where voters reward politicians for good performance beyond what they would in a standard prospective voting model. Previous work has also considered voters who use adaptive rules (Bendor et al. 2011), experience cognitive dissonance (Acharya, Blackwell and Sen 2018), focus more or less on costs/benefits (Nunnari and Zápal 2017), have partisan affect (Diermeier and Li 2019), selectively perceive new information (Gerber and Green 1999), or make random mistakes in their belief formation (Ogden 2016).[10]

2 A model of elections with motivated reasoners

We study a model of elections with voters who are directional motivated reasoners. The players are an incumbent politician (I, pronoun "she), a non-strategic electoral challenger (C), and a finite set of citizens $N = \{1, \ldots, n\}$ with n odd. We refer to a generic voter with j and male pronouns.

The incumbent politician has a competence θ_I. The value of θ_I is unknown to all players,

[9] Some influential applications from outside political science include studies of self-control (Bénabou and Tirole 2002) and investment (Brunnermeier and Parker 2005). More recently, Lipnowski and Mathevet (2017) study how to best persuade someone who forms beliefs in a manner similar to our voters.

[10] See also Levy and Razin (2015) for a model of information aggregation where voters don't account for the correlation between their sources of information. Outside of the electoral context, behavioral models have also been used to study the implications of citizens who are credulous about what politicians say (Little 2017). More generally, Minozzi (2013) considers a model of endogenous beliefs in which beliefs are chosen at the beginning of the game to maximize the total payoff from the game.

including the incumbent. The common prior belief is that θ_I is drawn from a normal distribution with mean μ_I and variance σ_θ^2. We normalize the prior expectation of incumbent competence so that $\mu_I = 0$. The challenger has analogous competence θ_C, drawn from a common prior distribution which is normal with mean μ_C and variance σ_θ^2.[11] The incumbent exerts effort $e \geq 0$, which is valued by the public but is not observed directly. This effort term captures a variety of things that an incumbent politician can choose to do to improve government performance indicators. Accountability is 'working well' when the incumbent exerts high levels of effort. The cost of effort is $c(e)$, which we assume is increasing and convex.

Though voters do not observe the incumbent's competence or effort directly, they observe a signal correlated with both. In particular, voters observe a public signal,

$$s = \theta_I + e + \varepsilon,$$

where ε is normally distributed with mean 0 and variance σ_ε^2. A natural way to interpret s is as a performance indicator like GDP growth. The ε term could correspond to factors outside of the incumbent's control that nonetheless also affect the outcome. Alternatively, ε could represent noise in citizens' perception of this signal – in which case $\theta_I + e$ is the "real" outcome. After the public signal is realized, there are also common shocks to voters' utilities for reelecting the incumbent (η_I) or instead electing the challenger (η_C). These shocks capture swings in candidate- or party-specific preferences that are unrelated to performance. We assume $\eta_C - \eta_I$ follows a continuous distribution F with support on the real line.

Each voter j also has an *affinity* for the incumbent, denoted by $a_j \in \mathbb{R}$, which directly affects his utility. Voter affinities capture reasons that voters' general taste ($a_j > 0$) or distaste ($a_j < 0$) for the incumbent that is independent of actual performance. Prominent examples include partisanship and party polarization (Druckman, Peterson and Slothuus 2013), party identification as social identity (Huddy, Mason and Aarøe 2015), or, more generally, motivated cognitive processes

[11] The variance of the belief about the challenger does not play a role in our analysis, so setting it equal to the variance of the incumbent belief is innocuous.

leading to beliefs based on loyalty to important affinity groups (Kahan 2012).[12] As the magnitude of an individual voter's affinity for (against) the incumbent grows, his motivation to conclude that the incumbent is highly competent (incompetent) increases, independent of performance. Accordingly, voters with different affinities may form different conclusions about incumbent competence, as we describe in more detail below.

Finally, after observing s, η_I, and η_C each citizen $j \in N$ decides whether to vote for the incumbent ($r_j = 1$) or the challenger ($r_j = 0$). The candidate receiving the majority of votes wins. We denote whether the incumbent is re-elected by $R \in \{0,1\}$, where $R = 1$ means the incumbent is reelected and $R = 0$ means the challenger is elected. To simplify the analysis we assume that if the incumbent wins, each voter's second-term payoff consists of the incumbent's competence, his affinity for the incumbent, and the random preference shock associated with the incumbent ($\theta_I + a_j + \eta_I$). If instead the challenger wins, each voter receives a second-term payoff equal to the challenger's competence and the random preference shock associated with the challenger ($\theta_C + \eta_C$).[13]

Payoffs. The incumbent seeks reelection but also pays for any effort invested in performance. Accordingly, her utility is given by the following expression:

$$u_I(e;R) = R - c(e).$$

In exerting higher effort, the incumbent trades off improving the chances for a high performance signal, which improves voter assessments of her quality, with the cost of that effort.

The utility of each voter $j \in N$ is given by,

$$u_j(R) = \underbrace{s + a_j}_{\text{Period 1}} + \underbrace{R(\theta_I + a_j + \eta_I)}_{\text{Period 2 with } I} + \underbrace{(1-R)(\theta_C + \eta_C)}_{\text{Period 2 with } C}.$$

[12] Affinities can also capture emotional states that may affect decision-making such as anxiety (Albertson and Gadarian 2015) or anger (Phoenix 2019, Wayne 2019, Webster 2018).

[13] These payoffs capture what the average performance in period 2 would be with no incumbent effort, which is what would happen in equilibrium since the game ends after second period effort choice.

The $s + a_j$ terms capture citizen payoffs from the first period, which are equal to the outcome s plus his affinity for the incumbent.[14] The remaining terms capture the second-period payoff based on incumbent and challenger behavior described above. Notice that $a_j < 0$ captures scenarios in which voter j has a general distaste for being represented by the incumbent, while $a_j > 0$ denotes a setting where the voter values being represented by the incumbent for reasons unrelated to performance. These affinities play a central role in our conceptualization of motivated reasoning.

Incorporating motivated reasoning. To capture citizens as motivated reasoners, we assume that rather than forming beliefs about incumbent competence using only Bayes's rule, as in standard accountability models, the voter forms a *conclusion* about incumbent competence. We refer to a generic conclusion as $\tilde{\theta}_I$, and assume voters reach an *optimal conclusion* $\tilde{\theta}_I^*$:

$$\tilde{\theta}_I^*(s; a_j) \in \arg\max_{\tilde{\theta}_I} \log f_{\theta_I|s}(\tilde{\theta}_I|s) + \delta v(a_j, \tilde{\theta}_I), \tag{1}$$

where the first term (the log-likelihood of the Bayesian posterior belief) captures the accuracy motive and the second term captures the directional motive (see Little 2019, for further discussion of this formulation).

In terms of the accuracy motive, by standard calculations, when the voter expects effort level $\hat{e}$ the first term is normally distributed with mean:

$$\overline{\mu}(s) = \frac{\sigma_\varepsilon^2(s - \hat{e}) + \sigma_\theta^2 \mu_I}{\sigma_\varepsilon^2 + \sigma_\theta^2},$$

and variance:

$$\overline{\sigma}_\theta^2 = \frac{\sigma_\varepsilon^2 \sigma_\theta^2}{\sigma_\varepsilon^2 + \sigma_\theta^2}.$$

That is, $\overline{\mu}(s)$ and $\overline{\sigma}_\theta^2$ represent the mean and variance of voters' Bayesian posteriors given perfor-

[14] The voter's utility from the first period does not affect any players' decisions in the model but plays a role in the welfare analysis later. For this portion of the utility, it is better to think of the noise term as factors affecting performance outside of the incumbent's ability or effort rather than noise in citizen perception. The equilibrium analysis is identical if we replace s with $\theta + e$, and the welfare analysis is the same on average.

mance signal s. The second term, $\delta v(a_j, \tilde{\theta}_I)$, measures each voter's directional motive. The $\delta \geq 0$ term is a scalar that measures the general strength of directional motivated reasoning, and the $v(\cdot)$ function dictates the relationship between the voter affinity and preferred conclusions. We assume that this function has the following properties.

Assumption 1. The directional motive function $v(a_j, \tilde{\theta}_I)$ is (i) continuous and (weakly) concave in $\tilde{\theta}_I$; and (ii) $\frac{\partial^2 v(a_j, \tilde{\theta}_I)}{\partial a_j \partial \tilde{\theta}_I} \geq 0$.

The continuity assumption in part (i) just means that small changes in the voter affinity and conclusion have small effects on the desirability of the conclusion. The concavity assumption loosely means that there are "diminishing returns" to forming a conclusion closer to one's ideal. Part (ii) states that as a voter's affinity for the incumbent increases, he has an intrinsic reason, aside from objective performance, to conclude that the incumbent is more competent. This ensures that directional motivations are increasing in voter affinity for the incumbent.

In effect, we make two departures from a standard model of belief formation, though only one of them is consequential in our setting. First, rather than computing an expected payoff given a belief about θ_I, the voter picks a single conclusion about the incumbent's competence. That is, even if $\delta = 0$, the voters here form a conclusion at the mode of their belief rather than caring about the full distribution. However, since θ_I enters the voter utility in a linear fashion and the mean and mode of a normal distribution are equal, a voter with $\delta = 0$ behaves in an identical manner to a voter using a standard Bayesian belief and expected utility maximization. The second, consequential difference is the addition of the directional motive, which we illustrate through a series of substantive examples.

Examples of directional motives. Before analyzing the general model we provide examples of directional motives that capture various conceptualizations of motivated reasoning.

Example 1 (*Polarized partisanship*)**.** It is natural to conceptualize citizens with positive affinity for the incumbent, $a_j > 0$, as the incumbent's co-partisans and those with negative incumbent affinity, $a_j < 0$, as citizens from the opposite party. Under this conceptualization directional motives repre-

sent the idea that co-partisans are motivated to believe that the incumbent is competent and those in the other party are motivated to believe that the incumbent is incompetent. Furthermore, those who more strongly identify with their partisan label (as represented by the magnitude of $|a_j|$) are more strongly motivated. This can be represented by the linear directional motive $v(a_j, \tilde{\theta}_I) = a_j \tilde{\theta}_I$. So, if voter j has affinity $a_j > 0$ ($a_j < 0$) then he is motivated to conclude that the incumbent is highly competent (incompetent). With this assumption, the optimal conclusion maximizes,

$$\log f_{\theta_I|s}(\tilde{\theta}_I|s) + \delta v(a_j, \tilde{\theta}_I) = \kappa - \frac{(\tilde{\theta}_I - \overline{\mu}(s))^2}{2\overline{\sigma}_\theta^2} + \delta a_j \tilde{\theta}_I,$$

where $\kappa = -\log(\overline{\sigma}_\theta^2) - \frac{1}{2}\log(2\pi)$ is a constant that does not depend on $\tilde{\theta}_I$. This is maximized at,

$$\tilde{\theta}_I^{\text{part}}(s, a_j) = \overline{\mu}(s) + \delta \overline{\sigma}_\theta^2 a_j, \tag{2}$$

In this case, voter conclusions approach the mean of the Bayesian posterior ($\tilde{\theta}_I^{\text{part}}(s, a_j) \to \overline{\mu}(s)$) under three conditions. First is if the directional motive becomes very weak $\delta \to 0$. Second is if voter affinity towards the incumbent is neutral $a_j \to 0$. Both of these follow from the fact that the v term approaches zero, and the maximum of a normal density is at the mean. The third condition is if the Bayesian belief is very precise $\overline{\sigma}_\theta \to 0$, which leads to very steep losses in the accuracy motive for even small deviations from the Bayesian mean.[15] Conversely, a motivated reasoner of this form will tend to form a belief far from the Bayesian mean when δ, $|a_j|$, and $\overline{\sigma}_I$ are large.

Example 2 (*Confirmation bias*)**.** Another type of motivation unrelated to affinity for the incumbent is a reluctance to process information in order to confirm preexisting beliefs.[16] Under the common prior assumption, this can be represented by $v(a_j, \tilde{\theta}_I) = -(\tilde{\theta}_I - \mu_I)^2$.[17] In this case, voter j is

[15] This last observation is consistent with recent experimental research arguing that weaker prior beliefs about a candidate allows for more effective persuasion by campaigns attempting to shift voter beliefs about candidate desirability (Broockman and Kalla 2020).

[16] See Rabin and Schrag (1999) for a microfoundation for this kind of belief formation. See Lockwood (2017) for a related study of confirmation bias in a political agency model with a focus on pandering dynamics.

[17] This is also related to the model of cognitive dissonance in Acharya, Blackwell and Sen (2018).

motivated to conclude that his initial assessment of incumbent competence was accurate. He still forms a new conjecture about competence through processing new information, but he is biased toward confirming his prior. Under this directional motive the first-order condition for voter j's optimal conclusion becomes,

$$-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta 2(\mu_I - \tilde{\theta}_I) = 0,$$

which is solved by,

$$\tilde{\theta}_I^{\text{conf}}(s, a_j) = \frac{1}{1+2\delta\overline{\sigma}_\theta^2}\overline{\mu}(s) + \frac{2\delta\overline{\sigma}_\theta^2}{1+2\delta\overline{\sigma}_\theta^2}\mu_I. \tag{3}$$

The confirmation bias directional motive creates a common bias for all citizens toward the prior that does not depend on a_j. Instead it takes the form of a weighted average of a fully Bayesian conclusion, $\overline{\mu}(s)$, and the (common) prior expectation of incumbent competence, μ_I. Similar to the polarized partisanship motive above the optimal conclusion approaches that of a Bayesian as directional motives approach zero, $\delta \to 0$, and the precision of the Bayesian belief increases, $\overline{\sigma}_\theta \to 0$. However, in this case voter j's directional motives push him to confirm him preexisting assessment of the incumbent rather than forming highly positive or negative assessments, as above, and therefore his affinity for the incumbent has no impact on distortions away from the Bayesian posterior mean. The manner in which voters form optimal conclusions in this example are analogous to the description of "Cautious Bayesians" in Hill (2017). Conclusions are formed in the correct direction of the information received, due to weight on voter accuracy motives, but except under particular circumstances the conclusions will not reach those of a perfect Bayesian, given a positive directional motive.

Example 3 (*Spatial motivations*)**.** Another interpretation of directional motives, which ends up combining elements of the first two, is that there is be a one-to-one correspondence between a citizen's affinity for the incumbent and his most preferred conclusion about the incumbent's competence. This can be represented by the quadratic directional motive $v(a_j, \tilde{\theta}_I) = -(a_j - \tilde{\theta}_I)^2$. In

this case, voter j is motivated to form a conclusion about incumbent quality that justifies his preexisting affinity for the incumbent. In that sense, his affinity is like an 'ideal conclusion target,' much like an ideal point operates in standard spatial models. The directional motive in this case takes the same general form as Example 2 above, leading to a similar form of the the optimal conclusion,

$$\tilde{\theta}_I^{\text{spat}}(s,a_j) = \frac{1}{1+2\delta\overline{\sigma}_\theta^2}\overline{\mu}(s) + \frac{2\delta\overline{\sigma}_\theta^2}{1+2\delta\overline{\sigma}_\theta^2}a_j. \tag{4}$$

As with confirmation bias example, this can be thought of as a weighted average of what a Bayesian would think ($\overline{\mu}(s)$) and what a "pure motivated reasoner" ($\delta \to \infty$) would think (a_j). Also similar to the previous examples, as the directional motive weakens, $\delta \to 0$, or precision of the Bayesian belief increases, $\overline{\sigma}_\theta \to 0$, the optimal conclusion approaches the Bayesian mean. However, an important consequence of spatial directional motives that runs counter to the polarized partisanship example is that for a voter with a neutral affinity (a_j close to zero), the spatially motivated belief does *not* approximate the Bayesian mean (as long as $\overline{\mu}(s) \neq a_j$). This is because a centrist who does not strongly support or oppose the incumbent *does not lack a directional motive*, but has a directional motive to form a conclusion close to zero. In other words, they are *motivated moderates* who intrinsically like holding a neutral view of the incumbent. As we will see when analyzing the voting model, this has important consequences for the politician's effort incentives.

As these examples illustrate, our model of motivated reasoning is very flexible and can incorporate a variety of voter biases. Citizens' directional motives have two principal effects on the relationship between a politician's performance and citizens' assessment of the politician. First, motivated reasoning may lead to *divergence*, meaning that voters' conclusions about politician quality differ based on their existing affinities for the incumbent. The second potential effect from directional motivated reasoning is *desensitization*, which is a weakening of the relationship between incumbent performance and citizens' beliefs.

The two effects of directional motivated reasoning are evident in the main examples from the fact that each optimal conclusion is decomposed into two components. One component is a "slope"

Directional Motive Example	Belief Divergence	Belief Desensitization
Polarized Partisanship	✓	X
Confirmation Bias	X	✓
Spatial Motivations	✓	✓

Table 1: Properties of different directional motives

term that multiplies the correct posterior mean/mode $\overline{\mu}(s)$: this represents the potential desensitization effects. In the partisan polarization example, the posterior mode is multiplied by one, so there is no desensitization effect in that case. In the confirmation bias and spatial motivations examples, the posterior mode is multiplied by $\frac{1}{1+2\delta\overline{\sigma}_\theta^2} \in (0,1)$, which indicates some level of desensitization in the sense that conclusions will depend less on the signal than under full Bayesian updating. Another component of each solution is an "intercept" term, which does not change the dependence of conclusions on the signal but does shift the baseline conclusion up or down. In Examples 1 and 3 this intercept term is strictly increasing in a_j, which indicates a divergence effect. Table 1 displays the properties of the directional motives in Examples 1-3 and Figure 1 illustrates these three directional motives.

Each panel of Figure 1 plots the optimal conclusion for three voters as a function of the signal of incumbent performance. One voter has a positive affinity towards the incumbent ($a_j = .3$, black lines), one a negative affinity towards the incumbent ($a_j = -3$, light grey lines), and one is neutral ($a_j = 0$, dark grey lines). The dotted lines correspond to a case with a relatively weak directional motive ($\delta = 1$), and the solid lines to a relatively strong directional motive ($\delta = 5$). The vertical line correspond to the average (expected) signal.

The left panel illustrates the polarized partisan motive. In this case, all of the lines (between voters and across levels of motivated reasoning) are parallel, indicating no desensitization. However, comparing across the dotted to solid lines, we see there is divergence as the solid lines are further apart from each other than the dotted lines. Another important aspect of this graph is that the centrist voter reaches the same conclusions (which correspond to the Bayesian mean) regard-

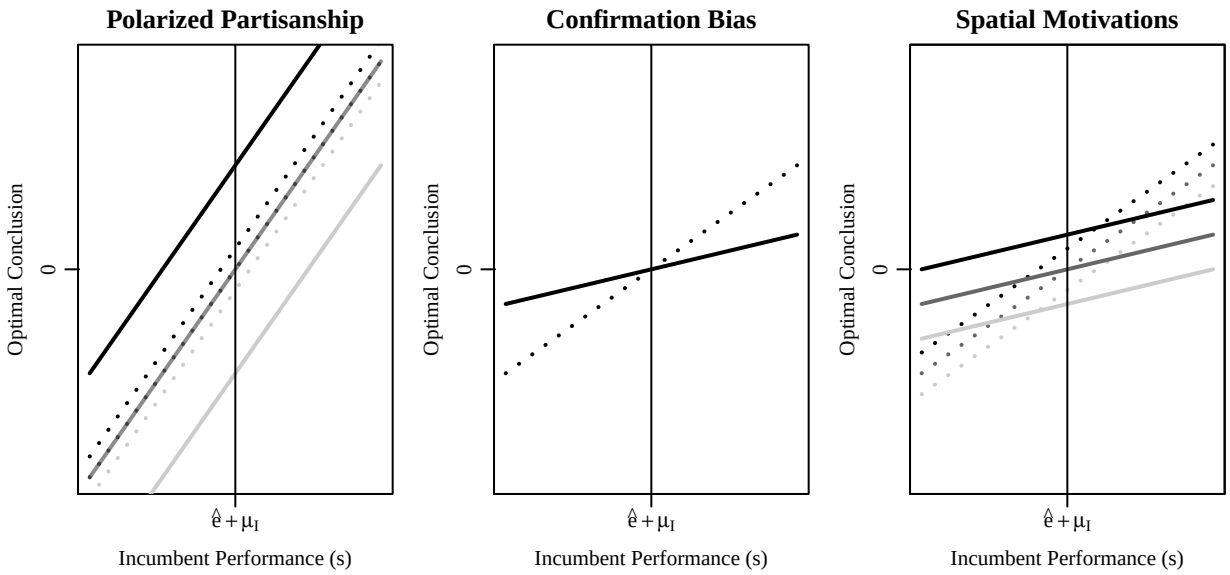

Figure 1: Illustration of the three main examples of directional motives.

less of the level of motivated reasoning.

The middle panel illustrates the confirmation bias directional motive. Here there is just one line because all of the voters reach the same conclusion independent of a_j. However, increasing the strength of the directional motive affects all voters, as this decreases the slope of the relationship between performance and conclusion. Thus, this motive illustrates a case of pure belief desensitization without divergence.

Finally, the right panel illustrates the spatial motivation. As strength of directional motive increases (dotted to solid lines), the slopes of the conclusions decrease, again indicating desensitization. However, in contrast to confirmation bias, there is also divergence in this case, which is easiest to see by looking at the vertical line (average signal), where the neutral voter reaches the same conclusion regardless of the directional motive, but when the directional motive increases the pro- and anti-incumbent voters move further apart.

In sum, for all three of these examples motivated reasoning is consistent with different voters responding to information in parallel (Guess and Coppock 2018), though it can change the slope of their updates (desensitization) and the difference between their intercept terms (divergence).

As we will argue below, motivated reasoning that leads to desensitization presents more serious

problems for politicians' responsiveness than does belief divergence. To make this argument, we first complete the description of the model with our solution concept.

Solution concept. A *sincere motivated reasoning equilibrium* (hereafter, "equilibrium") is a (pure) strategy profile consisting of an effort choice for the incumbent and a voting strategy for each voter $\left(e^*, \{r_j^*(s, \eta_I, \eta_C)\}_{j \in N}\right)$, combined with a profile of optimal conclusions for each voter, $\left\{\{\tilde{\theta}^*(s, a_j)\}_j \in N\right\}$ such that each conclusion $\tilde{\theta}^*(s, a_j)$ is a solution to Equation (1), each r_j assigns j's vote choice to j's most-preferred candidate under the conclusion $\tilde{\theta}(s, a_j)$ (we say $r_j^*(s, \eta_I, \eta_C) = 1$ if j votes for the incumbent and 0 otherwise),[18] and e^* maximizes u_I given the citizens' strategies.

3 Motivated reasoning and democratic accountability

Voter conclusions and voting behavior. We analyze the general model by first considering optimal conclusions and actions for the voters and then characterizing the incumbent politician's optimal effort level of as a function of voters' strategies. As above, we let $\hat{e} > 0$ denote the effort level that voters expect from the incumbent.

We introduce an additional assumption on v to guarantee the existence of an optimal conclusion, which holds for all of the examples described above.

Assumption 2. Either (i) $\frac{\partial v}{\partial \tilde{\theta}_I}$ is bounded above and below, or (ii) $\lim_{\tilde{\theta}_I \to \infty} \frac{\partial v}{\partial \tilde{\theta}_I} = -\infty$ and $\lim_{\tilde{\theta}_I \to -\infty} \frac{\partial v}{\partial \tilde{\theta}_I} = \infty$.

Assumption 2 rules out some extreme beliefs. Without this assumption we cannot rule out a situation in which higher (or lower) conclusions are always better for some voter and therefore a maximum is never reached. Lemma 1 establishes the existence of a unique optimal conclusion with intuitive comparative statics.

Lemma 1. *Under assumptions 1 and 2 there exists a unique optimal conclusion $\tilde{\theta}^*(s, a_j, \delta; \hat{e})$ for each voter $j \in N$. Furthermore, $\tilde{\theta}^*(s, a_j, \delta; \hat{e})$ is weakly increasing in a_j and s.*

[18] By a standard argument, voters must behave sincerely if we rule out weakly dominated strategies.

Hereafter we maintain assumptions 1 and 2. The role of Assumption 2 in Lemma 1 is only to establish existence: if Assumption 2 was violated but a solution happened to exist, the comparative statics results would still hold at that solution. One implication of the fact that $\tilde{\theta}^*$ is increasing in a_j is that voters' preferences satisfy single crossing: if voter j weakly prefers to retain the incumbent and $a_j < a_{j'}$ for some other voter j', then voter j' strictly prefers to retain the incumbent. Formally, assuming the voter retains the incumbent when indifferent, voter j votes for the incumbent if and only if:

$$\tilde{\theta}^*(s, a_j, \delta; \hat{e}) + a_j + \eta_I \geq \mu_C + \eta_C, \tag{5}$$

where the left-hand side of Equation (5) is strictly increasing in a_j. As a result, the voter with the median affinity is decisive. Let m be the index corresponding to the voter with median affinity so that a_m is the median voter's affinity for the incumbent and r_m^* is his equilibrium choice.[19]

Corollary 1. *In any equilibrium, the median voter is decisive:* $R = 1$ *if and only if* $\tilde{\theta}^*(s, a_m, \delta; \hat{e}) + a_m + \eta_I \geq \mu_C + \eta_C$.

For some of the results involving incumbent effort we will consider a special case of the model in which $\frac{\partial v}{\partial \tilde{\theta}_I}$ is linear in $\tilde{\theta}_I$ and a_j, as is true in all of our leading examples:

Assumption 3. (a) $\frac{\partial v}{\partial \tilde{\theta}_I}$ is linear in $\tilde{\theta}_I$ and (b) $\frac{\partial v}{\partial \tilde{\theta}_I}$ is linear in both $\tilde{\theta}_I$ and a_j.

Some of our results only rely on the weaker assumption 3a, while others are clearer with the stronger version 3b. Lemma 2 shows that these assumptions allow us to provide a relatively simple linear form to the optimal conclusion.

Lemma 2. *(i) Under assumption 3a, the optimal conclusion is linear in* $s - \hat{e}$*, i.e., it can be written:*

$$\tilde{\theta}^*(s, a_j, \delta; \hat{e}) = \alpha(a_j) + \beta(s - \hat{e}) \tag{6}$$

[19] A key assumption driving median responsiveness is that the shocks η_I and η_C affect all voters the same. In a more standard probabilistic voting model (see Coughlin 1992) where shocks are voter-specific, the voter with the median affinity will not necessarily be the effective median voter.

for some increasing function $\alpha(a_j)$ and $\beta \geq 0$ (which does not depend on a_j). β is strictly decreasing in δ if and only if v is strictly concave in $\tilde{\theta}_I$.

(ii) Further, if 3b holds then we can write the $\alpha(a_m)$ function in (6) as

$$\alpha(a_j) = \alpha_0 + \alpha_1 a_j \tag{7}$$

for some $\alpha_0 \in \mathbb{R}$ and $\alpha_1 \geq 0$. α_1 is strictly increasing in δ if and only if $\frac{\partial^2 v(a_j, \tilde{\theta}_I)}{\partial \tilde{\theta}_I \partial a_j} > 0$.

Lemma 2 gives us a convenient way to decompose the optimal conclusion into two components, which correspond exactly to the discussion of Figure 1. The first is an "intercept" term independent of the signal, $\alpha_0 + \alpha_1 a_j$, which depends more on the voter affinities (i.e., divergence effects are stronger) when α_1 is high. Belief divergence is stronger as the directional motive increases (higher δ) as long as those with strictly higher affinity for the incumbent want to form strictly higher conclusions. In our leading examples, this is true in the partisan and spatial motive, but not the cautious motive. The second is a "slope" term $\beta(s - \hat{e})$ which represents how sensitive a voter is to changes in observed incumbent performance. This term is strictly decreasing in δ – meaning more desensitization – if and only if the v function is strictly concave in the $\tilde{\theta}_I$. In our leading examples this is true with the spatial motivations and confirmation bias examples, but not the polarized partisanship example.

Incumbent effort. Our analysis in this section focuses on which aspects of directional motivated reasoning might lead to changes in incumbent behavior. To analyze which aspects of directional motivated reasoning matter for incumbent effort, we treat α_1 and β as parameters and perform comparative statics for the incumbent's decision. Further, since the median voter is decisive, we can analyze how these terms affect the optimal conclusion of the median voter (and hence effort).

What is the effect of divergence on effort? Formally, does increasing α_1 reduce the marginal return for the incumbent to work harder? The answer to this question is ambiguous. First, since the median is decisive, divergence effects only matter for politician behavior if the median is drawn to one side or the other. In some cases we may expect that the voter with median affinity, by virtue

of his position in the middle, is not strongly biased toward one candidate or another. Second, the median voter's bias for or against the incumbent mainly affects the marginal return on effort by making the election either closer or less close. For this reason, the effect of α_1 is ambiguous: if the incumbent is behind in the election ex ante, then increasing α_1 will reduce effort if the median voter dislikes the incumbent ($a_m < 0$) and increase effort if he likes the incumbent ($a_m > 0$). Conversely, if the incumbent is ahead ex ante, then increasing α_1 will tend to decrease effort when the median voter likes the incumbent ($a_m > 0$) and increase effort when he dislikes him ($a_m < 0$) because that increases electoral competitiveness. Overall, the effect of divergence on incumbent effort primarily depends on whether ex ante outcomes are pushed closer or further away from an even vote share.For this result we will say that the incumbent is behind if his ex ante probability of winning is less than one-half and otherwise he is ahead.

Proposition 1. *Under assumption 3b:*
(i) If $a_m = 0$, then divergence has no impact on incumbent effort.
(ii) If $a_m \neq 0$, then increasing divergence (i.e., increasing α_1) increases effort when either the incumbent is behind and $a_m > 0$ or the incumbent is ahead and $a_m < 0$, and decreases effort otherwise.

While this result shows that belief divergence can lead to more or less effort, there is a sense in which it will lead to less effort in more "typical" scenarios. In particular, the incumbent will tend to be electorally advantaged when the median voter has a positive affinity for her, and will tend to be disadvantaged otherwise. These are precisely the cases where further divergence will lead to less effort. Divergence only leads to more effort in "mismatched" cases where the median voter has a positive affinity for the incumbent but other factors (like being objectively less competent) make her disadvantaged, or vice versa. Further, there are limits to when stronger belief divergence can be good for accountability, because if motivated reasoning becomes strong enough, the incumbent will always be advantaged if and only the median voter has positive affinity towards him.

The effect of desensitization on incumbent effort is much more straightforward. Whenever belief desensitization increases (i.e., β increases) incumbent effort decreases.

Proposition 2. *Under assumption 3a, equilibrium incumbent effort is reduced by desensitization effects of motivated reasoning (e^* is increasing in β)*

Combined with Lemma 2, this implies that whenever the directional motive is strictly concave in $\tilde{\theta}_I$, stronger directional motives will lead to less effort through this channel. Combining Propositions 1 and 2, we can also state that when directional motives become sufficiently strong the overall effect of motivated reasoning (combining divergence and desensitization) must be to decrease incumbent effort, thereby harming accountability.

Corollary 2. *If $a_m \neq 0$ and $\frac{\partial^2 v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I \partial a_j} > 0$, then there exists a $\delta^* \geq 0$ such that incumbent effort is decreasing in δ for all $\delta \geq \delta^*$.*

To illustrate, Propositions 1 and 2 help us make some predictions regarding Examples 1-3. What effect would each type of directional motivated reasoning by voters have on politician behavior? Let us first consider the case in which $a_m = 0$ (i.e., the median voter has no clear preference for one candidate over the other). In such a case, the fact that voter conclusions diverge due to directional motives makes no difference for incumbent effort. In fact, from the incumbent's perspective, the polarized partisanship in Example 1 is no different at all from if the voters were fully Bayesian. However, Examples 2 and 3 also exhibit belief desensitization. In those cases the presence of directional motivated reasoning will still decrease incumbent effort since all voters', including the median's, optimal conclusions respond less to government performance information, which weakens incumbent effort incentives.

When $a_m \neq 0$ our predictions are more subtle. Figure 2 illustrates an example of how the partisan directional motive (Example 1) affects equilibrium effort and the incumbent's re-election probability. In each panel, the x-axis is the strength of the directional motive, so these plots illustrate how increasing the strength of voters' partisan motivations affects equilibrium outcomes. There are three cases: when the prior (Bayesian) belief is that the challenger is weaker than the incumbent ($\mu_C < \mu_I$; light grey), of equal competence on average ($\mu_C = \mu_I$; dark grey), or more competent than the incumbent ($\mu_C > \mu_I$; black).

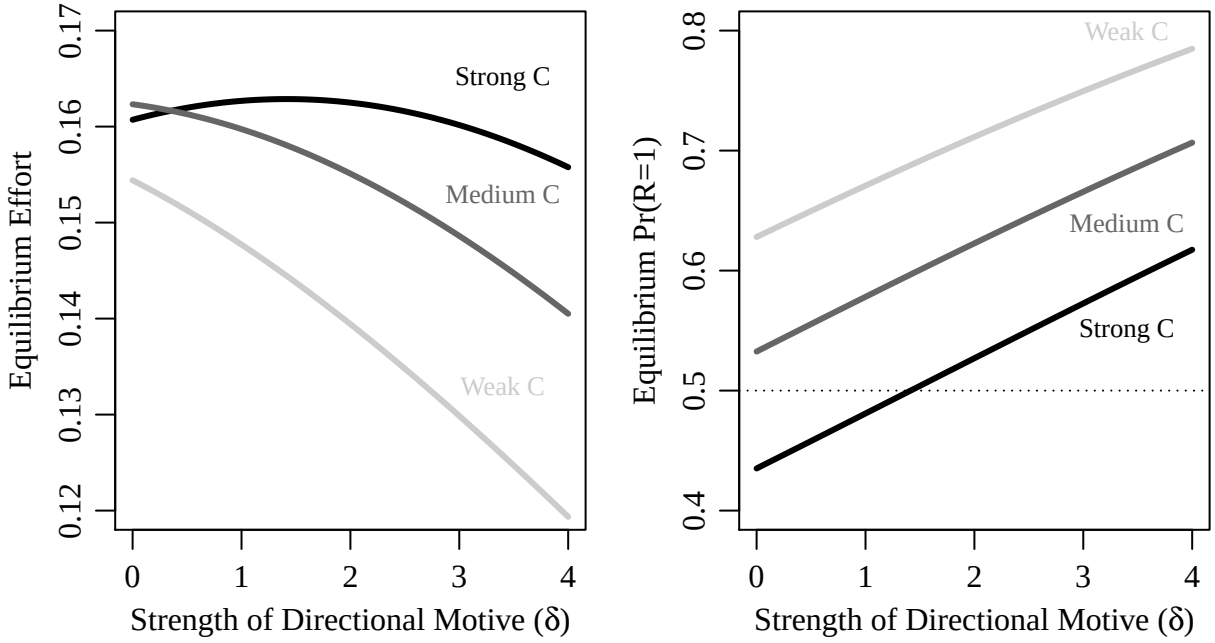

Figure 2: Equilibrium Effort and Re-election Probability with partisan directional motives, and a median voter with a small positive affinity for the incumbent

The left panel shows that if the challenger is believed to be weak or equally competent, stronger motivated reasoning always decreases incumbent effort. This is because, as we can see in the right panel, these are cases where the incumbent is likely to win. Since the median voter has a slight positive affinity for the incumbent, stronger directional motives make this advantage stronger, decreasing returns to effort.

Things are more interesting when the challenger is expected to be more competent than the incumbent. Here, starting in the right panel, we can see that with no directional motivated reasoning (low δ), the incumbent is more likely to lose the election. For this part of the parameter space, increasing directional motivated reasoning makes the election "closer", as the median voter affinity for the incumbent counteracts the fact that he is objectively less competent than the (expected) challenger. So, at low levels of δ, increasing directional motivated reasoning leads to *more* incumbent effort. However, as corollary 2 shows, when the directional motives become strong enough, eventually the incumbent is advantaged despite his relatively low competence and further increases in δ strengthen this advantage, which leads to decreases in equilibrium effort.

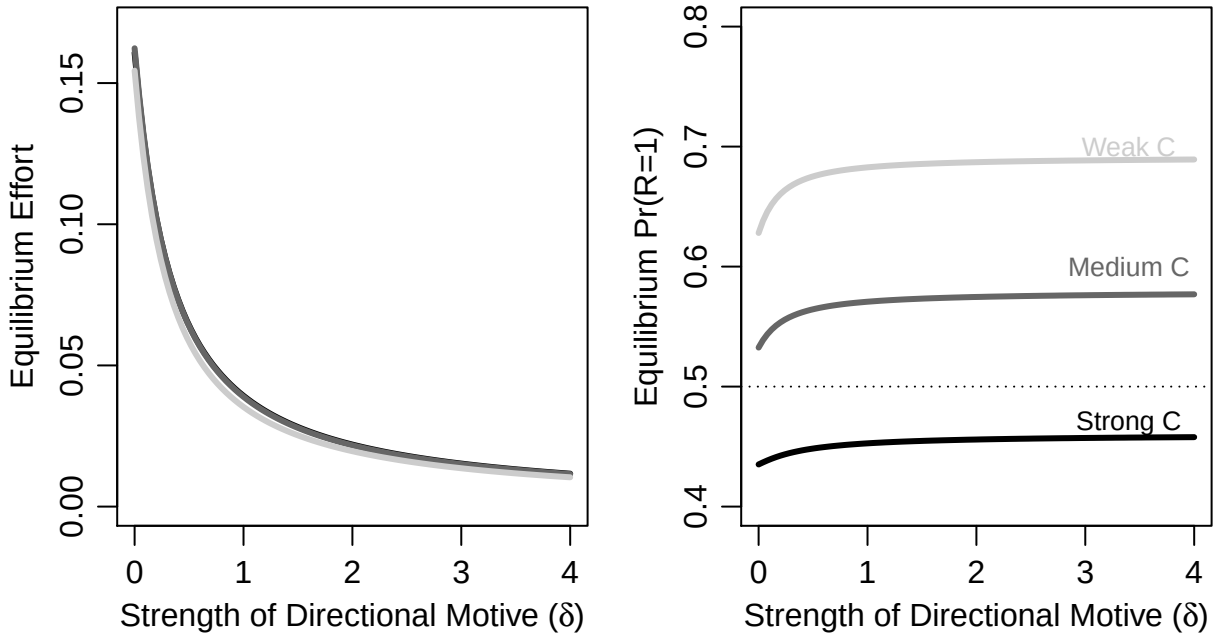

Figure 3: Equilibrium Effort and Re-election Probability with spatial directional motives, and a median voter with a small affinity for the incumbent

Figure 3 provides the same illustrations but using the spatial directional motive (Example 3), where there is desensitization in addition to divergence. Here the effect of increasing the directional motive on equilibrium effort is so strong (and, as expected, negative) that it swamps any differences based on how strong the challenger is. Though, as we can see in the right panel, there are still large differences across re-election probabilities.

Comparing across the figures, increased directional motivated reasoning that manifests primarily as desensitization leads to strong and unambiguous decreases in incumbent effort. On the other hand, increases in motivated reasoning which only affect belief divergence have generally weaker effects that can run in either direction.

3.1 Electoral selection and voter welfare

So far we have discussed the effects of motivated reasoning on politicians' incentives to invest in good governance but we have not discussed the overall effect of motivated reasoning on voter welfare. The total effect of motivated reasoning on voter welfare includes its effect on the Incumbent's effort level as well as its effect on the ability of voters to select good politicians. Consider the ex

ante welfare of the median voter:

$$W_m(e^*, r_m^*(s,\eta_I,\eta_C)) = \mathbb{E}_{(\theta,\varepsilon)}\Bigg[\theta + e^* + \varepsilon + \mathbb{E}_{(\eta_I,\eta_C)}\Big[r_m^*(\theta_I + e^* + \varepsilon, \eta_I, \eta_C)(\theta_I + a_j + \eta_I) + (1 - r_m^*(\theta_I + e^* + \varepsilon, \eta_I, \eta_C))(\mu_C + \eta_C)\Big]\Bigg]. \quad (8)$$

Motivated reasoning has two effects on the median voter's welfare. First, as we have already discussed, motivated reasoning may affect equilibrium effort, e^*. Second, motivated reasoning may affect second-period expected utility by changing the retention rule r_m^*, which depends on the optimal conclusion given δ. By Lemma 2, the optimal conclusion $\tilde{\theta}$ is linear in $s - \hat{e}$. Since $e^* = \hat{e}$ in equilibrium, this implies that the selection term in voter welfare is independent of effort. Thus, the effect of motivated reasoning decomposes the total effect of δ into two separable effects on accountability and selection.[20]

Examination of Equation (8) reveals that motivated reasoning must always have a negative effect on the selection component of voter welfare. When $\delta = 0$ the voters' decisions are identical to Bayesian decision-makers so voter welfare becomes,

$$\mathbb{E}_{(\theta,\varepsilon)}\left[\theta + e^* + \varepsilon + \mathbb{E}_{(\eta_I,\eta_C)}\left[\max\{\mathbb{E}[\theta_i|\theta_I + \varepsilon] + a_j + \eta_I, \mu_C + \eta_C\}\right]\right]. \quad (9)$$

In other words when $\delta = 0$ the median voter's retention strategy is the one that maximizes second-period expected utility given the available information because $\tilde{\theta}^*(s, a_j, \delta; \hat{e}) = \mathbb{E}[\theta_i|\theta_I + \varepsilon]$ (i.e. the optimal conclusion is the same as the correct conditional expectation). Adding motivated reasoning ($\delta > 0$) can only decrease second-period expected utility either by retaining the incumbent when $\mathbb{E}[\theta_i|\theta_I + \varepsilon] + a_j + \eta_I < \mu_C + \eta_C$ or removing him when $\mathbb{E}[\theta_i|\theta_I + \varepsilon] + a_j + \eta_I > \mu_C + \eta_C$.

Several conclusions follow about how motivated reasoning affects voter welfare. First, motivated reasoning has two effects on voter welfare: a strategic effect on effort by politicians and a statistical effect on selection of good politicians. Second, when motivated reasoning decreases

[20] This highlights a contrast between career concerns models of political agency like the one used here and signaling models. In a signaling model higher effort by low types reduces the effectiveness of selection by voters. Thus, in a signaling model higher effort would have more ambiguous welfare consequences than in our model.

effort by politicians it must always decrease voter welfare overall. Finally, when motivated reasoning increases effort by politicians it may still decrease voter welfare through its effect on selection, though this could be counterbalanced by positive strategic effects. Our interpretation of these results is that, though some counterexamples are possible, the presence of motivated reasoning is most likely to be bad news for voters.

4 The relationship between performance and vote shares

So far we have primarily studied how different kinds of directional motives affect accountability, or incumbent effort incentives. In our model this largely depends on the behavior of the median voter. However, a large body of empirical work on electoral accountability focuses on the relationship between aggregate vote shares — which depend on the behavior of all voters — and measures of incumbent success, most frequently economic performance (e.g., Erikson 1989, Fair 1996, Hall, Yoder and Karandikar 2017, Healy and Lenz 2017, Hopkins and Pettingill 2018, Kramer 1971, Lewis-Beck and Stegmaier 2000, Markus 1992, Tufte 1976).[21]

In this section we adapt our model to show that motivated reasoning can affect incumbent vote share without affecting electoral accountability in terms of incumbent effort incentives. To do so, we explore how different directional motives affect the relationship between expected incumbent vote share given government performance: $\mathbb{E}[VS|s]$. To begin, for a fixed signal of government performance s, voter j votes to retain the incumbent if,

$$Pr(\tilde{\theta}^*(s, a_j, \delta; \hat{e}) + a_j + \eta_I \geq \mu_C + \eta_C).$$

To straightforwardly convert this observation into expected vote shares we assume Assumption 3b holds and we assume that $\eta_I - \eta_C$ be normally distributed with mean μ_η and variance σ^2_η. Further, we assume that the electorate is sufficiently "large" and that voter affinities are normally distributed

[21] See Healy and Malhotra (2013), Margalit (2019), and Warshaw (2019) for useful reviews.

with mean μ_η and variance σ^2_η.[22] Under these assumptions,[23] we can write the condition for j to vote for the incumbent as,

$$\alpha_0 + \alpha_1 a_j + \beta_j(s - e^*) + a_j + (\eta_I - \eta_C) \geq \mu_C. \tag{10}$$

The left-hand side of Inequality (10) (again, for a fixed realization of s) is normally distributed with mean $\alpha_0 + \beta_j(s-e^*) + (1+\alpha_1)\mu_a + \mu_\eta$ and variance $(1+\alpha_1)^2\sigma_a^2 + \sigma_\eta^2$, so the incumbent's average vote share is given by,

$$\mathbb{E}[VS|s] = \Phi\left(\frac{\alpha_0 + \beta_j(s-e^*) + (1+\alpha_1)\mu_a + \mu_\eta - \mu_C}{\sqrt{(1+\alpha_1)^2\sigma_a^2 + \sigma_\eta^2}}\right) \tag{11}$$

As noted above, many empirical papers studying electoral accountability examine the relationship between incumbent vote shares and some performance indicator (in our model, s) such as change in GDP, change in real income, unemployment, crime, etc. In our formulation the theoretical prediction for this relationship is given by differentiating $\mathbb{E}[VS|s]$ with respect to s, which we denote with $\Delta^{VS}(s)$:

$$\Delta^{VS}(s) = \frac{\partial \mathbb{E}[VS|s]}{\partial s} = \frac{\beta_j}{\sqrt{(1+\alpha_1)^2\sigma_a^2 + \sigma_\eta^2}}\phi\left(\frac{\alpha_0 + \beta_j(s-e^*) + (1+\alpha_1)\mu_a + \mu_\eta - \mu_C}{\sqrt{(1+\alpha_1)^2\sigma_a^2 + \sigma_\eta^2}}\right) \tag{12}$$

When $\Delta^{VS}(s) = 0$ there is no relationship between observed government performance (s) and incumbent vote share ($\mathbb{E}[VS|s]$). When $\Delta^{VS}(s) \neq 0$ performance information does impact the incumbent's average vote share. The next result establishes that even when the effects of motivated reasoning do not impact equilibrium effort (e.g., there is no desensitization) there can nonetheless be an effect on vote shares.

[22] The preceding analysis holds for any finite number of voters; what matters here is that n is large enough that the number of voters with an affinity up to a_j is well-approximated by the *ex ante* probability that an individual voter has an affinity in this range.

[23] From the fact that in equilibrium $s = \theta_I + e^* + \varepsilon$ and $\hat{e} = e^*$ we could further simplify the optimal conclusion, but to connect more directly to the relevant empirical findings we want to write this as a function of the performance signal s.

Remark 1. *Suppose Assumption 3b holds, that* $(\eta_I - \eta_C)$ *is normally distributed with mean* μ_η *and variance* σ^2_η*, and voter affinities are normally distributed with mean* μ_a *and variance* σ^2_a*. Further, let* $a_m = 0$*,* $\mu_a = 0$*, and* $\frac{\partial \beta}{\partial \delta} = 0$ *(so that there is belief divergence but no desensitization, as in Example 1). Then there can be effects on incumbent vote share even though there are none on equilibrium effort.*

The preceding analysis showed that when the median voter is exactly centrist and motivated reasoning only causes divergence in voter conclusions, there is no change in incumbent effort. However, this does not mean divergence does not affect the relationship between vote share and performance. For example, consider the polarized partisan directional motive in Example 1. In this case, as the strength of the directional motive grows arbitrarily large ($\delta \to \infty$), $\alpha_1 \to \infty$, which leads to $\Delta^{VS}(s) \to 0$. This implies that the relationship between performance s and incumbent vote share diminishes as directional motives dominate voter conclusions. Moreover, given part (i) of Proposition 1, this implies that even when there is no effect of directional motives on incumbent effort there can still be an effect on the relationship between performance and vote shares. Intuitively, this is because, letting $a_m = 0$, when the belief divergence effect is very strong nearly all voters with positive affinity for the incumbent $a_j > 0$ vote to retain him while nearly all voters with negative affinity for the incumbent $a_j < 0$ vote to remove him, yet the median voter remains decisive for both the electoral outcome and for providing incumbent effort incentives. More generally, motivated reasoning producing belief divergence may change the relationship between incumbent performance and *vote shares*[24] even if it does not affect incentives for effort.

Figure 4 provides illustrative examples. The top panels show (similar to some of the previous diagrams) the relationship between the strength of the directional motive (δ) and effort choice of the incumbent for different model specifications. Each panel contains three vertical lines corresponding to three levels of directional motive, and each corresponding bottom panel plots the relationship between incumbent performance s and average vote share at these three levels of directional motive. In each panel, the solid line/curve corresponds to the case with no directional mo-

[24]Donovan et al. (2019) and Freeder (2019) provide evidence that this relationship has been weakening in the US.

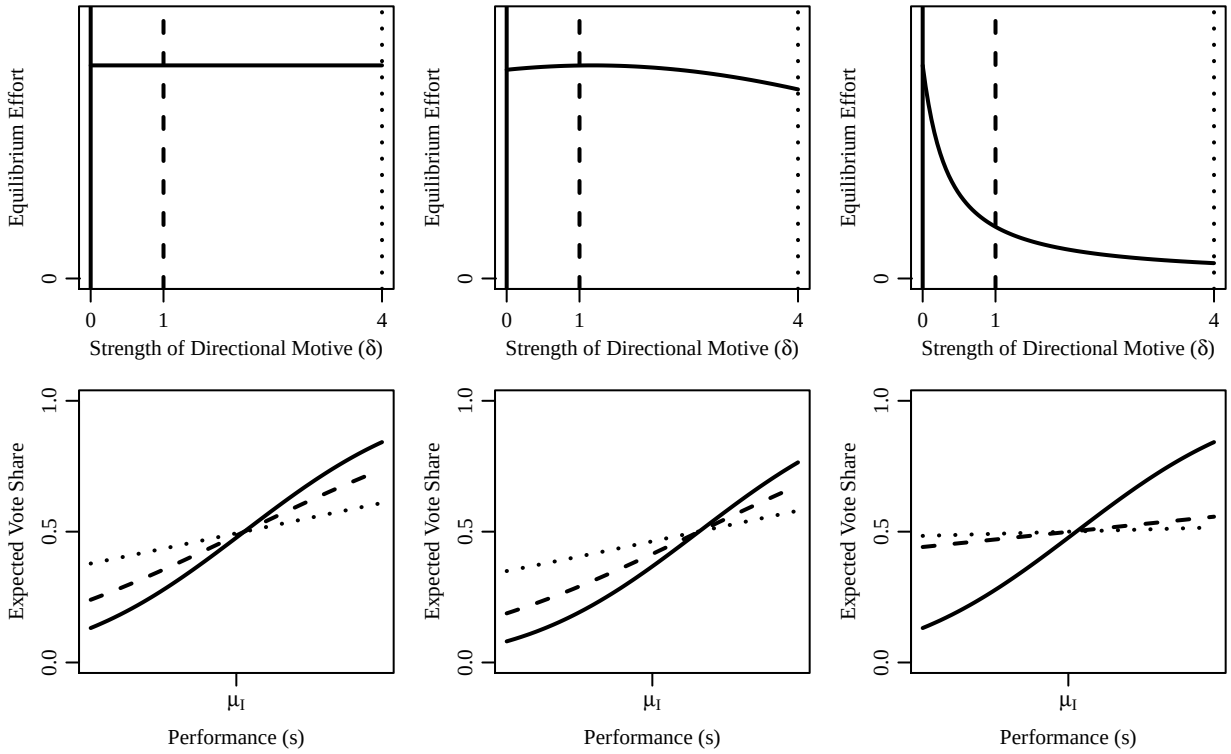

Figure 4: Relationship between vote share and performance for varying levels of motivated reasoning (bottom panels), with corresponding equilibrium effort choices (top panels).

tives ($\delta = 0$), the dashed line/curve to "low" directional motives ($\delta = 1$), and the dotted line/curve to "high" directional motives ($\delta = 4$).

The left-most panels represent a partisan directional motive as in Example 1 with a median voter is an exact centrist, $a_m = 0$. In this case variation in directional motive δ has no impact on effort incentives, as discussed above. However, the bottom-left panel shows that increasing the directional motive (solid to dashed to dotted curve) weakens the relationship between performance and vote share. This captures the driving environment underlying Remark 1: there is no desensitization effect since it is the partisan motive from Example 1, but there is belief divergence which does not affect equilibrium effort levels when $a_m = 0$ but it increasingly weakens the relationship between performance and incumbent vote shares as directional motivations intensify. Thus, it is possible that the relationship between performance and vote share can become weaker *while having no effect on accountability*.

The middle panel also models the partisan motive but now the median voter has a slight positive

affinity for the incumbent ($a_m = .15$). The challenger is also *ex ante* more likely to be competent ($\mu_I = 0,\ \mu_C = .4$) and therefore enjoys a competence-based electoral advantage when there are weak directional motives. As the analysis in Section 3 shows, increasing δ, which intensifies the median's positive affinity for the incumbent, 'tightens' the election by reducing the challenger's advantage. This translates into an initial increase in effort (as δ moves from 0 to 1). However, looking now at the corresponding bottom panel, the relationship between vote shares and performance is weaker moving from $\delta = 0$ (solid curve) to $\delta = 1$ (dashed curve). This illustrates that the connection between incumbent success and vote shares can be harmed even while there is a positive impact on effort incentives. Once δ increases sufficiently the median's positive affinity pushes the incumbent 'ahead', thereby reducing effort incentives. Intuitively, in this case, a weaker relationship between performance and vote shares corresponds to weakened effort incentives. This can be seen by comparing either $\delta = 0$ to $\delta = 4$ (dotted curve) or $\delta = 1$ to $\delta = 4$. It was worth noting, however, that while the slope of the dotted line is relatively flat compared to either the dashed or solid curves, it corresponds to about 90% of the effort incentives. In this case at least, very different relationships between performance and vote share are indicative of relatively modest changes in accountability.

Finally, the right panels represent a case with the spatial directional motive with an exact centrist median voter ($a_m = 0$). This is a case then where motivated reasoning leads to both belief divergence and desensitization. Accordingly, we see that effort decrease rapidly in directional motive (δ). In this case the relationship between performance and incumbent vote share is also strongly affected moving from $\delta = 0$ to either $\delta = 1$ or $\delta = 4$, with only modest different between the two latter levels of directional motive. Intuitively, as directional motives intensify, the dual impact of divergence and desensitization leads to weak accountability and a weak relationship between incumbent performance and vote shares. This example provides further illustration that the presence of desensitization effects from motivated reasoning can prove to be very detrimental for democratic accountability from multiple angles.

A weakening relationship between incumbent performance and vote shares might be a symp-

tom of weakening democratic accountability, but the degree to which this is true depends on what exactly drives this attenuation. If driven by desensitization, then a weakening relationship between incumbent performance and voter shares is cause for alarm. However, if it is mostly driven by belief divergence then this is not necessarily the case. Further, since effort can sometimes increase in belief divergence, it is possible to see a weaker performance-vote share relationship even while the incumbent faces stronger incentives to produce good performance.

5 Conclusion: A Behavioral/Rationalist Equivalence

We have shown different ways in which motivated reasoning might harm democratic accountability. But an important question is whether these are unique to our model in which voters trade-off accuracy and satisfaction of directional motivations. A benefit of reducing the total effect of motivated reasoning to two channels – belief divergence and belief desensitization – is that we can map any realization of our model with voters who are motivated reasoners to a "standard" model where voters process information only using Bayes's rule.

In particular, take any set of parameters to the model meeting assumption 3b where $\delta > 0$ and voters have affinities $\{a_j\}$. In this starting model, the voters choose to re-elect if and only if:

$$(\alpha_0 + \alpha_1 a_j + \beta(s - \hat{e})) + a_j \geq \mu_C \tag{13}$$

Now consider a modified version of the model where voters form their beliefs about θ_I using Bayes' rule. Keep all other parameters fixed, other than adding a parameter w which scales how much the voter cares about the incumbent performance (relative to the affinity), and let the voters have a different set of affinities a'_j. By a standard analysis, voters in such a model will vote to re-elect if and only if

$$w(\alpha_0 + \beta'(s - \hat{e})) + a'_j \geq \mu_C. \tag{14}$$

where α_0 is the same as in our main model and $\beta' = \frac{\sigma_\varepsilon^2}{\sigma_{\hat{\varepsilon}}^2 + \sigma_{\hat{\theta}}^2}$. If we set:

$$w = \beta / \beta' \tag{15}$$

$$a'_j = a_j(1 + \alpha_1) \tag{16}$$

then the behavior of all voters is identical in this modified game (for any fixed conjecture $\hat{e}$), which implies the incumbent has the exact same incentives and hence will choose the same effort level. That is, for any version of our model with motivated reasoning, there is a modified version of the model with fully Bayesian voters where (1) voters put less weight on the incumbent performance, and (2) have more spread out affinities, and in this modified version of the game the equilibrium behavior is identical.[25] It is possible to get a similar results by result by changing the variances of the prior belief – loosely, how informed voters are about politics – and noise term ε – loosely speaking, how much attention voters pay to the signal of performance or how much control politicians have over performance.[26]

In other words, while motivated reasoning certainly matters for democratic accountability, the effects are not qualitatively distinct from reformulating voter preferences in more standard models. As directional motives strengthen in our motivated reasoning model, voters behave as if they are more polarized in terms of preferences and/or place less emphasis, or pay less attention to, objective performance indicators when reaching conclusions about incumbent desirability. These general dynamics can also manifest in a standard accountability model through changes to the preference environment. Thus, many of the comparative statics from canonical models of electoral accountability mimic the effects of non-Bayesian belief formation present in our model.

Ultimately, we offer qualified agreement with the argument that voter biases that arise through directional motivated reasoning are bad news for democracy. Our results suggest that we should

[25]Formally, since we have shown that in the main model that β is decreasing in δ and α_1 is increasing in δ, we know that $w < 1$ in the modified game. So, for any voters k and l such that $a_k \neq a_l$, $|a'_k - a'_l| > |a_k - a_l|$.

[26]This equivalence is weaker in the sense that different priors entail changes to the distribution of the incumbent performance signals and hence the distribution of voter behavior across realizations of the game (if not for a particular realization of performance).

expect some aspects of directional motivated reasoning to have negative consequences for democratic accountability. However, the findings most often seen as the primary subject of concern are that partisans of different stripes hold different beliefs on statements of fact. This type of belief divergence does not seem to have a uniform, consistent effect on politicians' performance incentives. The larger concern, we argue, is the extent to which motivated reasoning may weaken the relationship between a politician's performance and conclusions about that politician. This may be present, for example, in the "cautious Bayesian" updating found in experimental work (Hill 2017) and need not have anything to do with partisanship. If motivated reasoning is a problem for democracy, the corrective may be to help voters become more sensitive to new information rather than focusing on reducing partisanship.

References

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2018. "Explaining Preferences from Behavior: A Cognitive Dissonance Approach." *The Journal of Politics* 80(2):400–411.

Achen, Christopher H. and Larry M. Bartels. 2017. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton University Press.

Adida, Claire, Jessica Gottlieb, Eric Kramon, Gwyneth McClendon et al. 2017. "Reducing or reinforcing in-group preferences? An experiment on information and ethnic voting." *Quarterly Journal of Political Science* 12(4):437–477.

Albertson, Bethany and Shana Kushner Gadarian. 2015. *Anxious politics: Democratic citizenship in a threatening world*. New York, NY: Cambridge University Press.

Alt, James, Ethan Bueno de Mesquita and Shanna Rose. 2011. "Disentangling accountability and competence in elections: evidence from US term limits." *The Journal of Politics* 73(1):171–186.

Anderson, Christopher J. 2007. "The end of economic voting? Contingency dilemmas and the limits of democratic accountability." *Annual Review of Political Science* 10:271–296.

Ashworth, Scott. 2005. "Reputational Dynamics and Political Careers." *Journal of Law, Economics, & Organization* 21(2):441–466.

Ashworth, Scott. 2012. "Electoral accountability: recent theoretical and empirical work." *Annual Review of Political Science* 15:183–201.

Ashworth, Scott and Anthony Fowler. 2019. "Electorates vs. Voters." *Working Paper* .
URL: *https://drive.google.com/file/d/1aMx-W24wzx19tEqwhjYUQVxiOyM9bVoh/view*

Ashworth, Scott and Ethan Bueno de Mesquita. 2008. "Electoral selection, strategic challenger entry, and the incumbency advantage." *The Journal of Politics* 70(4):1006–1025.

Ashworth, Scott and Ethan Bueno do Mesquita. 2014. "Is Voter Competence Good for Voters?: Information, Rationality, and Democratic Performance." *American Political Science Review* 108(3):565–587.

Bartels, Larry M. 2002. "Beyond the Running Tally: Partisan Bias in Political Perceptions." *Political Behavior* 24(2):117–150.

Bénabou, Roland and Jean Tirole. 2002. "Self-confidence and personal motivation." *The Quarterly Journal of Economics* 117(3):871–915.

Bendor, Jonathan, Daniel Diermeier, David A. Siegel and Michael M. Ting. 2011. *A Behavioral Theory of Elections*. Princeton University Press.
URL: *http://www.jstor.org/stable/j.ctt7shf3*

Bisgaard, Martin. 2019. "How getting the facts right can fuel partisan-motivated reasoning." *American Journal of Political Science* 63(4):824–839.

Bolsen, Toby, James N Druckman and Fay Lomax Cook. 2014. "The influence of partisan motivated reasoning on public opinion." *Political Behavior* 36(2):235–262.

Broockman, David E. and Joshua L. Kalla. 2020. "When and Why Are Campaigns' Persuasive Effects Small? Evidence from the 2020 U.S. Presidential Election." *Unpublished manuscript. University of California, Berkeley.* .
URL: *https://osf.io/m7326/*

Brunnermeier, Markus K and Jonathan A Parker. 2005. "Optimal expectations." *The American Economic Review* 95(4):1092–1118.

Bullock, John G, Alan S Gerber, Seth J Hill and Gregory A Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10(4):519–578.

Bullock, John G and Gabriel Lenz. 2019. "Partisan bias in surveys." *Annual Review of Political Science* 22:325–342.

Canes-Wrone, Brandice, Michael C Herron and Kenneth W Shotts. 2001. "Leadership and pandering: A theory of executive policymaking." *American Journal of Political Science* pp. 532–550.

Casella, George and Roger L. Berger. 2002. *Statistical Inference*. Thomson Learning.

Coughlin, Peter J. 1992. *Probabilistic Voting Theory*. Cambridge University Press.

De Benedictis-Kessner, Justin and Christopher Warshaw. N.d. "Accountability for the Local Economy at All Levels of Government in United States Elections." *American Political Science Review*. Forthcoming.

De Vries, Catherine E., Sara B. Hobolt and James Tilley. 2018. "Facing up to the facts: What causes economic perceptions?" *Electoral Studies* 51:115 – 122.
URL: *http://www.sciencedirect.com/science/article/pii/S0261379416304619*

Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.

Diermeier, Daniel and Christopher Li. 2017. "Electoral control with behavioral voters." *The Journal of Politics* 79(3):890–902.

Diermeier, Daniel and Christopher Li. 2019. "Partisan Affect and Elite Polarization." *American Political Science Review* 113(1):277–281.

Donovan, Kathleen, Paul M. Kellstedt, Ellen M. Key and Matthew J. Lebo. 2019. "Motivated Reasoning, Public Opinion, and Presidential Approval." *Political Behavior* .
URL: *https://doi.org/10.1007/s11109-019-09539-8*

Druckman, James N, Erik Peterson and Rune Slothuus. 2013. "How elite partisan polarization affects public opinion formation." *American Political Science Review* 107(1):57–79.

Druckman, James N and Mary C McGrath. 2019. "The evidence for motivated reasoning in climate change preference formation." *Nature Climate Change* 9(2):111–119.

Duggan, John and César Martinelli. 2017. "The Political Economy of Dynamic Elections: Accountability, Commitment, and Responsiveness." *Journal of Economic Literature* 55(3):916–84.
URL: *http://www.aeaweb.org/articles?id=10.1257/jel.20150927*

Eggers, Andrew C et al. 2014. "Partisanship and electoral accountability: Evidence from the UK expenses scandal." *Quarterly Journal of Political Science* 9(4):441–472.

Erikson, Robert S. 1989. "Economic Conditions and The Presidential Vote." *The American Political Science Review* 83(2):567–573.
URL: *http://www.jstor.org/stable/1962406*

Fair, Ray C. 1978. "The effect of economic events on votes for president." *The review of economics and statistics* pp. 159–173.

Fair, Ray C. 1996. "The effect of economic events on votes for president: 1992 update." *Political Behavior* 18(2):119–139.

Fearon, James D. 1999. Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance. In *Democracy, accountability, and representation*, ed. Adam Przeworski, Susan C Stokes and Bernard Manin. Cambridge University Press.

Ferraz, Claudio and Frederico Finan. 2008. "Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes." *The Quarterly journal of economics* 123(2):703–745.

Fowler, Anthony. 2020. "Partisan Intoxication or Policy Voting?" *Quarterly Journal of Political Science* 15(2):141–179.

Fowler, Anthony et al. 2016. "What Explains Incumbent Success? Disentangling Selection on Party, Selection on Candidate Characteristics, and Office-Holding Benefits." *Quarterly Journal of Political Science* 11(3):313–338.

Fox, Justin. 2007. "Government transparency and policymaking." *Public choice* 131(1-2):23–44.

Fox, Justin and Stuart V Jordan. 2011. "Delegation and accountability." *The Journal of Politics* 73(3):831–844.

Freeder, Sean. 2019. "ItâĂŹs No Longer the Economy, Stupid: Selective Perception and Attribution of Economic Outcomes." Unpublished Manuscript. University of California, Berkeley.

Gerber, Alan and Donald Green. 1999. "Misperceptions about perceptual bias." *Annual review of political science* 2(1):189–210.

Gordon, Sanford C, Gregory A Huber and Dimitri Landa. 2007. "Challenger entry and voter learning." *American Political Science Review* 101(2):303–320.

Green, Donald P, Bradley Palmquist and Eric Schickler. 2004. *Partisan hearts and minds: Political parties and the social identities of voters*. New Haven, CT: Yale University Press.

Guess, Andrew and Alexander Coppock. 2018. "Does Counter-Attitudinal Information Cause Backlash? Results From Three Large Survey Experiments." *British Journal of Political Science* pp. 1–19.

Hall, Andrew B. 2015. "What happens when extremists win primaries?" *American Political Science Review* 109(1):18–42.

Hall, Andrew B., Jesse Yoder and Nishant Karandikar. 2017. "Economic Distress and Voting: Evidence from the Subprime Mortgage Crisis." *Unpublished manuscript. Stanford University* . **URL:** *http://www.andrewbenjaminhall.com/HKY_foreclosures.pdf*

Healy, Andrew and Gabriel S. Lenz. 2017. "Presidential Voting and the Local Economy: Evidence from Two Population-Based Data Sets." *The Journal of Politics* 79(4):1419–1432. **URL:** *https://doi.org/10.1086/692785*

Healy, Andrew J, Neil Malhotra and Cecilia Hyunjung Mo. 2010. "Irrelevant events affect voters' evaluations of government performance." *Proceedings of the National Academy of Sciences* 107(29):12804–12809.

Healy, Andrew and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16:285–306.

Hill, Seth J. 2017. "Learning Together Slowly: Bayesian Learning about Political Facts." *The Journal of Politics* 79(4):1403–1418.

Hirano, Shigeo, Gabriel S Lenz, Maksim Pinkovskiy and James M Snyder Jr. 2015. "Voter learning in state primary elections." *American Journal of Political Science* 59(1):91–108.

Hopkins, Daniel J and Lindsay M Pettingill. 2018. "Retrospective voting in big-city US mayoral elections." *Political Science Research and Methods* 6(4):697–714.

Huddy, Leonie, Lilliana Mason and Lene Aarøe. 2015. "Expressive partisanship: Campaign involvement, political emotion, and partisan identity." *American Political Science Review* 109(1):1–17.

Jacobson, Gary C. 2010. "Perception, Memory, and Partisan Polarization on the Iraq War." *Political Science Quarterly* 125(1):31–56.

Kahan, Dan M. 2012. "Ideology, motivated reasoning, and cognitive reflection: An experimental study." *Judgment and Decision making* 8(4):407–24.

Kramer, Gerald H. 1971. "Short-term fluctuations in US voting behavior, 1896–1964." *American political science review* 65(1):131–143.

Kuklinski, James H., Paul J. Quirk, Jennifer Jerit, David Schwieder and Robert F. Rich. 2000. "Misinformation and the Currency of Democratic Citizenship." *The Journal of Politics* 62(3):790–816.

Kunda, Ziva. 1990. "The case for motivated reasoning." *Psychological bulletin* 108(3):480.

Lau, Richard R. and David P. Redlawsk. 2006. *How voters decide: Information processing in election campaigns*. New York, NY: Cambridge University Press.

Lenz, Gabriel S. 2013. *Follow the leader?: how voters respond to politicians' policies and performance*. New York, NY: University of Chicago Press.

Levy, Gilat and Ronny Razin. 2015. "Correlation Neglect, Voting Behavior, and Information Aggregation." *American Economic Review* 105(4):1634–1645.

Lewis-Beck, Michael S and Mary Stegmaier. 2000. "Economic determinants of electoral outcomes." *Annual review of political science* 3(1):183–219.

Li, Christopher, Gregory Sasso and Ian Turner. 2020. "Accountability in governing hierarchies." *Unpublished manuscript. Yale University* .

Lipnowski, Elliot and Laurent Mathevet. 2017. "Disclosure to a Psychological Audience." Manuscript.

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

Little, Andrew T. 2019. "The Distortion of Related Beliefs." *American Journal of Political Science* 63(3):675–689.

Lockwood, Ben. 2017. "Confirmation bias and electoral accountability.".

Lodge, Milton and Charles S Taber. 2013. *The rationalizing voter*. New York, NY: Cambridge University Press.

Lupia, Arthur and Mathew D. McCubbins. 1998. *The democratic dilemma: Can citizens learn what they need to know?* New York, NY: Cambridge University Press.

Margalit, Yotam. 2019. "Political responses to economic shocks." *Annual Review of Political Science* .

Markus, Gregory B. 1992. "The Impact of Personal and National Economic Conditions on Presidential Voting, 1956-1988." *American Journal of Political Science* 36(3):829–834.

Maskin, Eric and Jean Tirole. 2004. "The politician and the judge: Accountability in government." *American Economic Review* 94(4):1034–1054.

Milgrom, Paul and Chris Shannon. 1994. "Monotone Comparative Statics." *Econometrica* 62(1):157–180.

Minozzi, William. 2013. "Endogenous Beliefs in Models of Politics." *American Journal of Political Science* 57(3):566–581.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12021*

Nunnari, Salvatore and Jan Zápal. 2017. "A Model of Focusing in Political Choice." *CEPR Discussion Paper No. DP12407* .

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

Ogden, Benjamin. 2016. "The imperfect beliefs voting model." Manuscript. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2431447`.

Orr, Lilla V and Gregory A Huber. 2019. "The Policy Basis of Measured Partisan Animosity in the United States." *American Journal of Political Science* .

Payson, Julia A. 2017. "When Are Local Incumbents Held Accountable for Government Performance? Evidence from US School Districts." *Legislative Studies Quarterly* 42(3):421–448.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/lsq.12159*

Persson, T. and G.E. Tabellini. 2002. *Political Economics: Explaining Economic Policy*. Cambridge, MA: MIT Press.

Phoenix, Davin L. 2019. *The Anger Gap: How Race Shapes Emotion in Politics*. New York, NY: Cambridge University Press.

Popkin, Samuel L. 1991. "The Reasoning Voter: Communication and Persuasion in Presidential Campaigns.".

Prior, Markus, Gaurav Sood, Kabir Khanna et al. 2015. "You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions." *Quarterly Journal of Political Science* 10(4):489–518.

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

Redlawsk, David P. 2002. "Hot Cognition or Cool Consideration? Testing the Effects of Motivated Reasoning on Political Decision Making." *Journal of Politics* 64(4):1021–1044.

Redlawsk, David P, Andrew JW Civettini and Karen M Emmerson. 2010. "The affective tipping point: Do motivated reasoners ever "get it"?" *Political Psychology* 31(4):563–593.

Schnakenberg, Keith. 2020. "Candidate traits in elections: when good news for selection is bad news for accountability." *Political Science Research and Methods* pp. 1–9.

Strickland, April A, Charles S Taber and Milton Lodge. 2011. "Motivated reasoning and public opinion." *Journal of health politics, policy and law* 36(6):935–944.

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

Taber, Charles S. and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3):755–769.

Thaler, Michael. 2019. "The Fake News Effect : An Experiment on Motivated Reasoning and Trust in News." Manuscript.

Topkis, Donald M. 1978. "Minimizing a Submodular Function on a Lattice." *Operations Research* 26(2):305–321.

Tufte, Edward R. 1976. *Political Control of the Economy*. Princeton, NJ: Princeton University Press.

Warshaw, Christopher. 2019. "Local elections and representation in the United States." *Annual Review of Political Science* .

Wayne, Carly N. 2019. "The Goldilocks Problem of Counterterror: Constraints of an Emotional Electorate." *Unpublished manuscript. Washington University in St. Louis* .
URL: *https://drive.google.com/file/d/1NeKMD0N-Sgsrc9O0puYmP8cr24HcZKoC/view*

Webster, Steven W. 2018. "Anger and declining trust in government in the American electorate." *Political Behavior* 40(4):933–964.

Woon, Jonathan. 2012. "Democratic accountability and retrospective voting: A laboratory experiment." *American Journal of Political Science* 56(4):913–930.

Online Supplemental Appendix

A Main examples

In this section we provide the derivations for each of the three examples: polarized partisanship, spatial motivations, and confirmation bias. We also show when each optimal conclusion approaches what a perfect Bayesian would conclude as a function of directional motivation strength δ, voter affinity a_j, and the accuracy of the signal $\overline{\sigma}_\theta^2$ (as listed in Table 1). In general we have that each voter j forms an optimal conclusion by maximizing the following with respect to θ,

$$\begin{aligned}\log f_{\theta_I|s}(\tilde{\theta}_I|s) + \delta v(a_j,\tilde{\theta}_I) &= \log\left(\frac{1}{\sigma 2\pi} e^{-\frac{(\tilde{\theta}_I-\overline{\mu}(s)^2}{2\overline{\sigma}_\theta^2}}\right) + \delta v(a_j,\tilde{\theta}_I) \\ &= -\log(\sigma) - \frac{1}{2}\log(2\pi) - \frac{(\tilde{\theta}_I-\overline{\mu}(s))^2}{2\sigma^2} + \delta v(a_j,\tilde{\theta}_I).\end{aligned}$$

Differentiating yields the general first-order condition:

$$-\frac{\tilde{\theta}_I-\overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta\frac{\partial v(a_j,\tilde{\theta}_I)}{\partial\tilde{\theta}_I} = 0. \tag{17}$$

For each example we need only plug in the particular functional form for $v(a_j, \tilde{\theta}_I)$.

Polarized partisanship. In the first example in which voters are motivated to form 'large' conclusions (in absolute terms), in the direction of their affinities, we set $v(a_j, \tilde{\theta}) = \theta a_j$. Thus, $\frac{\partial v(a_j, \tilde{\theta}_I)}{\partial \tilde{\theta}_I} = a_j$. Plugging this into (17) we recover j's optimal conclusion from the first example:

$$-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}^2_\theta} + \delta a_j = 0,$$

$$\tilde{\theta}_I = \overline{\mu}(s) + \delta \overline{\sigma}^2_\theta a_j.$$

It is straightforward to see when the optimal conclusion approaches the mean of the Bayesian posterior:

$$\lim_{\delta \to 0} \left[\overline{\mu}(s) + \delta \overline{\sigma}^2_\theta a_j\right] = \overline{\mu}(s),$$

$$\lim_{\overline{\sigma}^2_\theta \to 0} \left[\overline{\mu}(s) + \delta \overline{\sigma}^2_\theta a_j\right] = \overline{\mu}(s),$$

$$\lim_{a_j \to 0} \left[\overline{\mu}(s) + \delta \overline{\sigma}^2_\theta a_j\right] = \overline{\mu}(s).$$

Confirmation bias. In the second example we set $v(a_j, \tilde{\theta}_I) = -(\theta_I - \mu_I)^2$ so voters are motivated to form conclusions near their prior. Thus, $\frac{\partial v(a_j, \tilde{\theta}_I)}{\partial \tilde{\theta}_I} = 2(\mu_I - \tilde{\theta}_I)$. This yields the first-order condition,

$$-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}^2_\theta} + \delta 2(\mu_I - \tilde{\theta}_I) = 0,$$

$$\tilde{\theta}_I = \frac{\overline{\mu}(s) + 2\delta \mu_I \overline{\sigma}^2_\theta}{1 + 2\delta \overline{\sigma}^2_\theta},$$

which can be rewritten, just as in the spatial motivations example, as,

$$\tilde{\theta}_I = \frac{1}{1 + 2\delta \overline{\sigma}^2_\theta} \overline{\mu}(s) + \frac{2\delta \overline{\sigma}^2_\theta}{1 + 2\delta \overline{\sigma}^2_\theta} \mu_I.$$

In terms of when the optimal conclusion approaches the fully Bayesian benchmark we have:

$$\lim_{\delta \to 0} \left[\frac{\overline{\mu}(s) + 2\delta \mu_I \overline{\sigma}_\theta^2}{1 + 2\delta \overline{\sigma}_\theta^2} \right] = \overline{\mu}(s),$$
$$\lim_{\overline{\sigma}_\theta^2 \to 0} \left[\frac{\overline{\mu}(s) + 2\delta \mu_I \overline{\sigma}_\theta^2}{1 + 2\delta \overline{\sigma}_\theta^2} \right] = \overline{\mu}(s),$$

and that clearly a_j does not impact distortions in this case.

Spatial motivations. In the final example where voters are motivated to match their conclusions to their affinity for the incumbent we set $v(a_j, \tilde{\theta}_I) = -(a_j - \tilde{\theta}_I)^2$. Accordingly, $\frac{\partial v(a_j, \tilde{\theta}_I)}{\partial \tilde{\theta}_I} = 2(a_j - \tilde{\theta}_I)$. Plugging in to (17) we have,

$$-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta 2(a_j - \tilde{\theta}_I) = 0,$$
$$\tilde{\theta}_I = \frac{\overline{\mu}(s) + 2\delta a_j \overline{\sigma}_\theta^2}{1 + 2\delta \overline{\sigma}_\theta^2},$$

which can be rewritten, analogous to the second example, as,

$$\tilde{\theta}_I = \frac{1}{1 + 2\delta \overline{\sigma}_\theta^2} \overline{\mu}(s) + \frac{2\delta \overline{\sigma}_\theta^2}{1 + 2\delta \overline{\sigma}_\theta^2} a_j.$$

We can characterize when the optimal conclusion approaches the mean of the Bayesian posterior as follows:

$$\lim_{\delta \to 0} \left[\frac{\overline{\mu}(s) + 2\delta a_j \overline{\sigma}_\theta^2}{1 + 2\delta \overline{\sigma}_\theta^2} \right] = \overline{\mu}(s),$$
$$\lim_{\overline{\sigma}_\theta^2 \to 0} \left[\frac{\overline{\mu}(s) + 2\delta a_j \overline{\sigma}_\theta^2}{1 + 2\delta \overline{\sigma}_\theta^2} \right] = \overline{\mu}(s),$$
$$\lim_{a_j \to 0} \left[\frac{\overline{\mu}(s) + 2\delta a_j \overline{\sigma}_\theta^2}{1 + 2\delta \overline{\sigma}_\theta^2} \right] = \frac{\overline{\mu}(s)}{1 + 2\delta \overline{\sigma}_\theta^2} \neq \overline{\mu}(s),$$

so we see that motivated reasoning still manifests in this case even when voter affinity a_j is zero.

B Proofs of results

Lemma 1. *Under assumptions 1 and 2 there exists a unique optimal conclusion* $\tilde{\theta}^*(s,a_j,\delta;\hat{e})$ *for each citizen* $j \in N$. *Furthermore,* $\tilde{\theta}^*(s,a_j,\delta;\hat{e})$ *is weakly increasing in* a_j *and in s.*

Proof of Lemma 1. The first-order condition for an optimal conclusion is

$$-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta \frac{\partial v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I} = 0. \tag{18}$$

Under Assumption 2, we have

$$\lim_{\theta\to\infty}\left[-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta \frac{\partial v(a_j,\tilde{\theta}_I)}{\partial \theta}\right] = -\infty$$

and

$$\lim_{\theta\to-\infty}\left[-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta \frac{\partial v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I}\right] = \infty.$$

Thus, we have $-\frac{\theta_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta \frac{\partial v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I} > 0$ for some $\tilde{\theta}_I < 0$ and $-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta \frac{\partial v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I} < 0$ for some $\theta > 0$. By continuity, there exists some $\tilde{\theta}^*(s,a_j,\delta;\hat{e})$ that solves 18. Strict concavity of the objective function implies that this is the unique maximum.

The fact that $\tilde{\theta}$ is increasing in a_j follows from the assumption that $\frac{\partial^2 v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I \partial a_j} \geq 0$. Applying the implicit function theorem to 18 gives

$$\frac{\partial \tilde{\theta}^*}{\partial a_j} = -\frac{\frac{\partial^2 v(a_j,\tilde{\theta}_I)}{\partial \theta \partial a_j}}{-\frac{1}{\overline{\sigma}^2_{\tilde{\theta}_I}} + \delta \frac{\partial^2 v}{\partial \tilde{\theta}_I^2}} > 0.$$

Hence, the optimal conclusion is increasing in a_j. Finally, the fact that $\tilde{\theta}^*$ is increasing in s follows from the fact that $\overline{\mu}(s)$ is increasing in s. Again applying the implicit function theorem, we have

$$\frac{\partial \tilde{\theta}^*}{\partial s} = -\frac{\frac{\overline{\mu}'(s)}{\overline{\sigma}^2_\theta}}{-\frac{1}{\overline{\sigma}^2_\theta} + \delta \frac{\partial^2 v}{\partial \tilde{\theta}_I^2}} > 0.$$

Thus, $\tilde{\theta}$ is also increasing in s. ■

Lemma 2. *(i) Under assumption 3a, the optimal conclusion is linear in $s-\hat{e}$, i.e., it can be written:*

$$\tilde{\theta}^*(s,a_j,\delta;\hat{e}) = \alpha(a_j) + \beta(s-\hat{e})$$

for some increasing function $\alpha(a_j)$ and $\beta \geq 0$ (which does not depend on a_j). β is strictly decreasing in δ if and only if v is strictly concave in $\tilde{\theta}_I$.
(ii) Further, if 3b holds then we can write the $\alpha(a_j)$ function in (6) as

$$\alpha(a_j) = \alpha_0 + \alpha_1 a_j$$

for some $\alpha_0 \in \mathbb{R}$ and $\alpha_1 \geq 0$. α_1 is strictly increasing in δ if and only if $\frac{\partial^2 v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I \partial a_j} > 0$.

Proof of Lemma 2. Recall that the first-order condition for an optimal conclusion is

$$-\frac{\tilde{\theta}_I - \overline{\mu}(s)}{\overline{\sigma}_\theta^2} + \delta \frac{\partial v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I} = 0.$$

With assumption 3a, we can write $\frac{\partial v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I} = \gamma_0(a_j) + \gamma_\theta(a_j)\tilde{\theta}_I$ for some γ_0 and γ_θ, which could both in principle be a function of a_j. Solving for $\tilde{\theta}_I$ yields

$$\tilde{\theta}_I(\cdot) = \frac{\gamma_0(a_j)\overline{\sigma}_\theta\delta}{1-\gamma_\theta(a_j)\overline{\sigma}_\theta\delta} + \frac{\overline{\mu}(s)}{1-\gamma_\theta(a_j)\overline{\sigma}_\theta\delta}.$$

Since $\overline{\mu}(s)$ is a linear function of $s-\hat{e}$ and $\overline{\sigma}_\theta$ is independent of the signal, this proves that $\tilde{\theta}(s,a_j,\delta;\hat{e})$ is linear in $s-\hat{e}$.

This also implies that $\gamma_\theta(a_j)$ must be constant in a_j; if not, the slope on the optimal conclusion with respect to s would be different for voters with different affinities, which would then contradict lemma 1, as for some $a_j'' > a_j'$ there would have to exist some signals where the optimal conclusion is higher for a_j'. Dropping the a_j argument from γ_θ and substituting in the formula for $\overline{\mu}(s)$, we

can write the optimal conclusion as:

$$\tilde{\theta}_I(\cdot) = \frac{\gamma_0(a_j)\overline{\sigma}_\theta\delta}{1-\gamma_\theta\overline{\sigma}_\theta\delta} + \frac{\frac{\sigma_\varepsilon^2(s-\hat{e})+\sigma_\theta^2\mu_I}{\sigma_\varepsilon^2+\sigma_\theta^2}}{1-\gamma_\theta\overline{\sigma}_\theta\delta}.$$

Setting:

$$\alpha(a_j) = \frac{\gamma_0(a_j)\overline{\sigma}_\theta\delta}{1-\gamma_\theta\overline{\sigma}_\theta\delta} + \frac{\frac{\sigma_\theta^2\mu_I}{\sigma_\varepsilon^2+\sigma_\theta^2}}{1-\gamma_\theta\overline{\sigma}_\theta\delta}$$

$$\beta = \frac{\frac{\sigma_\varepsilon^2(s-\hat{e})}{\sigma_\varepsilon^2+\sigma_\theta^2}}{1-\gamma_\theta\overline{\sigma}_\theta\delta}$$

completes part ii.

With assumption 3b, we can write $\frac{\partial v(a_j,\tilde{\theta}_I)}{\partial\tilde{\theta}_I} = \gamma_0 + \gamma_\theta\tilde{\theta}_I + \gamma_a a_j$ for some γ_0 (now a constant), γ_θ (which again can't be a function of a_j), and γ_a. Plugging this into the FOC and solving for $\tilde{\theta}_I$ yields

$$\tilde{\theta}_I = \frac{(\gamma_0+\gamma_a a_j)\overline{\sigma}_\theta\delta}{1-\gamma^\theta\overline{\sigma}_\theta\delta} + \frac{\overline{\mu}(s)}{1-\gamma^\theta\overline{\sigma}_\theta\delta}.$$

After performing similar substitutions as above, this is clearly linear and additively separable in a_j and $s-\hat{e}$. ■

Proposition 1. *Under assumption 3b:*

(i) If $a_m = 0$, then polarization has no impact on incumbent effort.

(ii) If $a_m \neq 0$, then increasing polarization (i.e. increases in α_1) increases effort when the incumbent is behind and $a_m > 0$ or the incumbent is ahead and $a_m < 0$, and decreases effort otherwise.

Proof of Proposition 1. As in the proof of Proposition 2, the marginal effect of effort on re-election is

$$\frac{1}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\int_0^1 \phi\left(\frac{\overline{s}(p;a_m,\alpha,\beta,\hat{e})-\mu_I-e}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\right)dp.$$

The second-order effect of α_m is

$$\frac{1}{\sigma_\theta^2+\sigma_\varepsilon^2}\int_0^1 -\frac{1}{\beta_m}\phi'\left(\frac{\bar{s}(p;a_m,\alpha,\beta,\hat{e})-\mu_I-e}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\right)dp.$$

Notice that the integrand positive when ϕ' is negative and negative when ϕ' is positive. If the incumbent is losing ex ante then she needs a better-than-expected signal to have an even chance at winning, so the negative part of this integral outweighs the positive part. If the incumbent is losing ex ante then she would need a worse-than-expected signal to have an even chance at winning, so the positive part of this integral outweighs the negative part. Combining these observations with the fact that α_m is increasing in α_1 if and only if $a_m > 0$ gives the result. ■

Proposition 2. *Under assumption 3a, equilibrium incumbent effort is reduced by desensitization effects of motivated reasoning (e^* is increasing in β).*

Proof of Proposition 2. Given Corollary 1 we can express the interim probability of re-election given a particular signal s as

$$\Pr[\text{re-election}|s] = \Pr[\eta_C-\eta_I \leq \tilde{\theta}^*(s,a_j,\delta;\hat{e})+a_m-\mu_C] = F(\tilde{\theta}^*(s,a_j,\delta;\hat{e})+a_m-\mu_C)$$

where F represents the distribution of $\eta_C-\eta_I$ (we said normal but I don't use that in this section). By Lemma 2, under assumption 3a we can set $\tilde{\theta}^*(s,a_j,\delta;\hat{e}) = \alpha_m+\beta_m(s-\hat{e})$ (where to reduce notation we write $\alpha(a_m)$ as α_m). We can invert this function for s to find, for any $t\in\mathbb{R}$, what signal would generate the conclusion t for voter m:

$$\tilde{\theta}_m^{-1}(t) = \frac{t+\beta_m\hat{e}-\alpha_m}{\beta_m}.$$

Since $F(\tilde{\theta}^*(s,a_j,\delta;\hat{e})+a_m-\mu_C)$ is strictly increasing in $\tilde{\theta}^*(s,a_j,\delta;\hat{e})$ it is also invertible, so $F^{-1}(p)-a_m+\mu_C$ tells us the conclusion that would lead to re-electing the incumbent with probability p. Putting these things together, we compute for each p the signal that would lead the

incumbent to be re-elected with probability p:

$$\overline{s}(p;a_m,\alpha_m,\beta_m,\hat{e}) = \tilde{\theta}_m^{-1}(F^{-1}(p)-a_m+\mu_C) = \frac{F^{-1}(p)-a_m+\mu_C-\alpha_m}{\beta_m}+\hat{e}.$$

Since s is normally distributed with mean μ_I+e and variance $\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}$, the distribution of the interim probability of re-election is therefore

$$\Phi\left(\frac{\overline{s}(p;a_m,\alpha_m,\beta_m,\hat{e})-\mu_I-e}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\right)$$

and, by the law of iterated expectations, the ex ante probability of re-election for a given effort level is simply the expectation of the interim probability. Using well-known results about expectations of nonnegative variables[27]

$$\Pr[\text{re-election}|e] = \int_0^1 \left[1-\Phi\left(\frac{\overline{s}(p;a_m,\alpha_m,\beta_m,\hat{e})-\mu_I-e}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\right)\right] dp.$$

Applying Liebniz's rule and using the chain rule inside of the integral, the marginal effect of effort on re-election is

$$\frac{\partial \Pr[\text{re-election}|e]}{\partial e} = \frac{1}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\int_0^1 \phi\left(\frac{\overline{s}(p;a_m,\alpha_m,\beta_m,\hat{e})-\mu_I-e}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\right) dp \tag{19}$$

and the second-order effect of changes in β_m is given by the derivative of (19) with respect to β_m:

$$\frac{\partial^2 \Pr[\text{re-election}|e]}{\partial e \partial \beta_m} = \frac{1}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\frac{1}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\int_0^1 \frac{\partial \overline{s}}{\partial \beta_m}\phi'\left(\frac{\overline{s}(p;a_m,\alpha_m,\beta_m,\hat{e})-\mu_I-e}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\right) dp \tag{20}$$

$$= \frac{1}{\sigma_\theta^2+\sigma_\varepsilon^2}\int_0^1 \frac{\alpha_m-F^{-1}(p)+a_m-\mu_C}{\beta_m^2}\phi'\left(\frac{\overline{s}(p;a_m,\alpha_m,\beta_m,\hat{e})-\mu_I-e}{\sqrt{\sigma_\theta^2+\sigma_\varepsilon^2}}\right) dp, \tag{21}$$

[27] See, for instance, Casella and Berger (2002) p. 78.

where the second expression involves plugging in our definition of $\bar{s}$ and differentiating with respect to β_m.

When this expression is positive we can say that desensitization reduces effort. We will show that the integrand is positive at every value of p. That is,

$$\frac{\alpha_m - F^{-1}(p) + a_m - \mu_C}{\beta_m^2} \phi' \left(\frac{\bar{s}(p; a_m, \alpha_m, \beta_m, \hat{e}) - \mu_I - e}{\sqrt{\sigma_\theta^2 + \sigma_\varepsilon^2}} \right) > 0$$

for all $p \in [0,1]$. Plugging our expression for $\bar{s}$ into the density term, this is

$$\frac{\alpha_m - F^{-1}(p) + a_m - \mu_C}{\beta_m^2} \phi' \left(\frac{\frac{F^{-1}(p) - a_m + \mu_C - \alpha_m}{\beta_m} + \hat{e} - \mu_I - e}{\sqrt{\sigma_\theta^2 + \sigma_\varepsilon^2}} \right) > 0.$$

In equilibrium we have $e = \hat{e}$ so we write

$$\frac{\alpha_m - F^{-1}(p) + a_m - \mu_C}{\beta^2} \phi' \left(\frac{\frac{F^{-1}(p) - a_m + \mu_C - \alpha_m}{\beta_m} - \mu_I}{\sqrt{\sigma_\theta^2 + \sigma_\varepsilon^2}} \right) > 0.$$

Intuitively, showing that this integrand is positive for all values of p is the same as showing that the two terms always have the same sign. The first term is positive for values of p such that:

$$F^{-1}(p) \geq \alpha_m + a_m - \mu_C. \tag{22}$$

Given our normalization that $\mu_I = 0$, when (22) holds the expression inside the ϕ^{-1} is negative, and since ϕ is single peaked at zero, the ϕ^{-1} term is also positive. Conversely, for the values of p where (22) does not hold and hence the first term is negative, the ϕ^{-1} term is also negative. Thus, increasing β increases marginal returns to effort in equilibrium. By the usual results about optima of functions with increasing differences (e.g. Topkis 1978, Milgrom and Shannon 1994), this implies that increasing β increases effort and decreasing β decreases effort. ■

Corollary 2. *If $a_m \neq 0$ and $\frac{\partial^2 v(a_j, \tilde{\theta}_I)}{\partial \tilde{\theta}_I \partial a_j} > 0$, then there exits a $\delta^* \geq 0$ such that incumbent effort is*

decreasing in δ for $\delta \geq \delta^*$.

Proof of corollary 2. If increasing δ decreases β, we know that effort decreases through this channel. So, it is sufficient to show that the result holds in the case where the only effect of increasing δ is through polarization (i.e., changing α_1).

If $\frac{\partial^2 v(a_j,\tilde{\theta}_I)}{\partial \tilde{\theta}_I \partial a_j} > 0$ then α_1 is strictly increasing in δ, and from inspecting the expressions in the proof of proposition 2 we can see it increases without bound. So, if $a_m > 0$, $\alpha_0 + \alpha_1 a_m$ approaches ∞ as $\delta \to \infty$, which ensures the incumbent is ahead. Conversely, if $a_m < 0$, as $\delta \to \infty$ the incumbent is sure to be behind. ■

Remark 1. *Suppose Assumption 3b holds, that $(\eta_I - \eta_C)$ is normally distributed with mean μ_η and variance σ_η^2, and voter affinities are normally distributed with mean μ_a and variance σ_a^2. Further, let $a_m = 0$, $\mu_a = 0$, and $\frac{\partial \beta}{\partial \delta} = 0$ (so that there is belief divergence but no desensitization, as in Example 1). Then there can be effects on incumbent vote share even though there are none on equilibrium effort.*

Proof of remark 1. Follows from argument/derivations in text given Proposition 1 showing that when $a_m = 0$ belief divergence does not affect effort and the fact that there is no belief desensitization effects when $\frac{\partial \beta}{\partial \delta} = 0$. ■