Essays on Information in Politics and Social Decisions

Inauguraldissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Carl Heese

aus Wesel, Nordrhein-Westfalen

Bonn, 2020

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Stephan Lauermann
Zweitreferent:	Prof. Dr. Daniel Krähmer
Tag der mündlichen Prüfung:	21. August 2020

Acknowledgments

This thesis would not have been possible without the support of many people. First of all, I would like to thank my advisors Stephan Lauermann and Daniel Krähmer. You have supported me greatly during the last years. The many discussions with you helped me develop my understanding of economics and my instinct about what relevant economics is. It is a pleasure to have you as supervisors. I thank Stephan Lauermann for his knowledgeable and always detailed advice on all the aspects of being an economist. I thank Daniel for his generosity and for being always supportive of my ideas. I also thank Daniel for making me part of the Collaborative Research Center Transregio 224, from which I benefited greatly.

I benefited greatly from the environment at the Bonn Graduate School of Economics. I am also thankful for support by a grant from the European Research Council (ERC 638115, Lauermann), and the support at the London School of Economics, and Yale University where I spent research stays during which part of this dissertation was completed. I would like to thank Philipp Strack for his supervision at Yale. I thank Thomas Tröger for his feedback and for being part of my dissertation committee. Special thanks go to Dezsö Szalay who helped me out with virtuosity in a crucial situation during the academic job market.

I am happy for many great moments with my fellow students and friends during my PhD. I thank Andreas Klümper for sharing the journey to Yale with me, and for all the hours of soccer and team sports. I thank Patrick Lahr for being such a nice roommate at Yale. I thank Gasper Ploj and Axel Wogrolly for many invigorating discussions about everybody and his dog. I thank my co-author and friend Deniz Kattwinkel; our time together at the Bonn Graduate School of Economics was always very inspiring and lead to a great academic collaboration. I thank Lucas Croé for being my Chinese learning partner and for many joyful meetings with good food and drinks, together with Xueying Liu. I thank Birgit Mauersberger and Rebecca Hader for their compassion and mindful gifts.

Most of all, I want to thank my family: my new Chinese family, in particular, Yulai Chen and Anna Li, who approach life with the most joyfulness, which makes me always wish that my Chinese would already be better. I want to thank my siblings Anne, Christiane, and Clemens, my grandfather Karl-Heinz—it makes me extremely happy to know that he can share these important moments of my life—, and my

parents Barbara Maria and Hans-Christian for their lifelong support. Finally, I want to thank my wife Si for being the wonderful person that she is. Thank you for being part of my life!

Contents

List of Figures

List of Tables

Introduction

This thesis consists of two parts. The first part consists of two chapters. The second part has one chapter.

How does the information of citizens shape the democratic process? This is the question asked in the first part. Each chapter of the first part proposes a specific economic model and analyzes a particular dimension of this question. **Chapter 1: "Persuasion and Information Aggregation in Elections"**, which is joint work with Stephan Lauermann, analyzes the scope of persuasion of voters by interested third parties. How manipulable are elections by third parties who hold private information and can strategically release relevant information to affect voters' behavior. Examples are numerous: in a shareholder vote, the management may strategically provide information regarding a potential merger through presentations and conversations; similarly, lobbyists provide selected information to legislators to influence their vote. We show that a manipulator can ensure that a majority of a large electorate supports his favorite policy simply by releasing some additional information to the voters. Moreover, persuasion does not require detailed knowledge about the citizens, the precise distribution of their preferences or their previous information. With very little knowledge about these, a third party manipulates by sending out private signals randomly to the citizens. A numerical example shows that persuasion is effective in elections with as few as 15 voters.

Chapter 2: "Voter Attention and Distributive Politics" studies how citizens paying attention to politics (or not) affects election outcomes, social welfare and its distribution. Demographic groups care differently much about different issues: e.g. older people care more about healthcare issues, while changes in education policy are more relevant to citizens with children. People that care more, pay more attention. We show that this attention effect shifts election outcomes into a direction that improves the overall welfare of a society. Elections often lead to outcomes that maximize a weighted welfare rule: the implicit decision weight of each voter is higher when he cares more about the issue voted on; however, less so when information is more cheap. In general, the decision weight is proportional to how informed the voter is. These results are important as they stress that information is a critical determinant of democratic participation. They imply that uninformed voters have

effectively almost no voting power, and that elections are susceptible to third-party manipulation of voter information.

Taken together, the first two chapters shed light upon two general topics. First, political actors seek to influence the citizens' opinions and behavior through propaganda, by the diverting of attention of the citizens, or by spreading false information, even more so in the digital age. How manipulable elections are through such informational tools? The first two chapters point out that the scope of manipulation is rather large. These insights may serve as a starting point for studying related questions, that, I believe, are highly relevant and deserve further analysis. The second broader topic this thesis touches upon is how the incentives of individuals shape their political beliefs. The second chapter points out that the size of the incentives matters since it affects the precision of people's beliefs, and thereby their implicit decision weight in elections. Studying the interaction of incentives and political beliefs has a positive motivation: we have a very limited understanding on how people form political beliefs, let alone why beliefs differ so much. But it also has a normative motivation since it informs about the consequences of economic interventions that shape the incentives.

The second part of the thesis is devoted to the social dimension of incentives and their role for belief formation, taking a step back from the political environment, however. Much empirical evidence has shown that many people depart from maximizing their self-interest, if doing so benefits others.[1] This means that these individuals' decisions are not sorely governed by their material desires, but also by "social motives". The recent research on *motivated reasoning* shows that many people deviate from complete egoism in order to 'feel moral' (for a review, see Gino, Norton, and Weber, 2016). It argues that, in social decisions, individuals can behave selfishly without a guilty conscience if they can make themselves *believe* that the selfish decision harms no others (for a review, see Gino, Norton, and Weber, 2016). In **Chapter 3: "Motivated Information Acquisition in Social Decisions"**, which is joint work with Si Chen, we ask: when do people stop acquiring information before a decision where pursuing one's own material benefits *might* harm others. Examples include medical examinations that help a doctor to decide between treatments with different profits, media consumption of voters before casting a ballot on ethically controversial policies, or consumers choosing to get informed about potential ethical issues of the products they would like to buy. Using a laboratory experiment, we provide causal evidence that having a selfishly preferred option makes individuals more likely to *continue* their inquiry for information when the information received up to that point suggests that the selfish behavior harms others. In contrast, when the information received up to that point suggests that being selfish harms nobody, indi-

1. For example, people donate to charity (e.g. DellaVigna, List, and Malmendier, 2012), pay postage to return misdirected letters (e.g. Franzen and Pointner, 2013), and share wealth with strangers in laboratory dictator games (e.g. Forsythe, Horowitz, Savin, and Sefton, 1994).

viduals are more likely to *stop* acquiring information. In some sense, individuals are fishing for excuses to behave selfishly until they find them. We also provide a theoretical model, drawing on the Bayesian Persuasion literature (Kamenica and Gentzkow (2011)). The model shows that the information acquisition strategy documented in our experiment can be optimal for a Bayesian agent who values the belief of herself not harming others but attempts to persuade herself to behave self-interestedly. Further, we empirically and theoretically provide results regarding the externalities that might not be obvious at first sight. Although one might think that strategic information acquisition must lead to more negative externalities when motivated by selfish interests, our model shows that also the reverse can happen: for some agent types, motivated information acquisition *improves* the welfare of the others affected by the decision. This counter-intuitive result rests on the observation that an "unmotivated" agent faces a moral hazard problem: when unmotivated, some agent types acquire only a small amount of information due to, for example, the satisficing behavior (Simon, 1955). The agent's selfish preference for one option over the other can mitigate this moral hazard problem by causing her to acquire more information in order to make sure that she chooses her least-preferred option only when certain that it is harmless to others. This result implies that delegating information acquisition to a neutral investigator might lower the welfare of the others affected by the decision.

References

DellaVigna, Stefano, John A List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56. [2]

Forsythe, Robert, Joel L Horowitz, Nathan E Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments." *Games and Economic behavior* 6 (3): 347–369. [2]

Franzen, Axel, and Sonja Pointner. 2013. "The External Validity of Giving in the Dictator Game." *Experimental Economics* 16 (2): 155–169. [2]

Gino, Francesca, Michael I Norton, and Roberto A Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212. [2]

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590–2615. [3]

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118. [3]

Chapter 1

Voter Persuasion and Information Aggregation in Elections

Joint with Stephan Lauermann

In most elections, a voter's ranking of outcomes depends on her information. For example, a shareholder's view of a proposed merger depends on her belief regarding its profitability and a legislator's support of proposed legislation depends on her belief regarding its effectiveness. An interested party that has private information may utilize this fact by strategically releasing information to affect voters' behavior. Examples of interested parties holding and strategically releasing relevant information for voters are numerous: in a shareholder vote, the management may strategically provide information regarding the merger through presentations and conversations; similarly, lobbyists provide selected information to legislators to influence their vote.

We are interested in the scope of such "persuasion" (Kamenica and Gentzkow, 2011) in elections. We study this question in the canonical voting setting by Feddersen and Pesendorfer (1997): there are two possible policies (outcomes)—*A* and *B*. Voters' preferences over policies are heterogeneous and depend on an unknown state, α or β, in a general way (some voters may prefer *A* in state α, some prefer *A* in state β, and some "partisans" may prefer one of the policies independently of the state). The preferences are drawn independently across voters and are each voters' private information. In addition, all voters privately receive information in the form of a noisy signal. The election determines the outcome by a simple majority rule.

In this setting, Feddersen and Pesendorfer (1997) have shown that within a broad class of "monotone" preferences and conditionally i.i.d. private signals, all equilibrium outcomes of large elections are equivalent to the outcome with a publicly known state ("information aggregation"). We restate their result as a benchmark in Theorem 1.

We ask the following question: can a manipulator ensure that a majority supports his favorite policy—potentially state-dependent—in a large election by providing *additional* information to the voters? Formally, the manipulator can choose

and commit to any joint distribution over states and signal realizations that are then privately observed by the voters. In particular, the manipulator's additional signal is required to be independent of the voters' exogenous private signals and their individual preferences (it is an "independent expansion"). The previous result by Feddersen and Pesendorfer (1997) suggests a limited scope for persuasion because, if voters simply ignored the additional information, the outcome would be "as if" the state were known, and, hence, the information provided by the manipulator would be worthless.

Our main result (Theorem 4) shows that, perhaps surprisingly, within the same class of monotone preferences and for any state-contingent policy, there exists an independent expansion of the voters' exogenous i.i.d. signal and an equilibrium that ensures that the targeted policy is supported by a majority with probability close to 1 when the number of voters is large. Thus, just by providing additional information, a manipulator can implement, for example, a targeted policy that is, in every state, the opposite of the outcome with full information.

The additional information affects the voters' behavior directly, by changing their beliefs about the state, and indirectly, by affecting their inference from being "pivotal" for the election outcome. While the direct effect is limited by the well-known "Bayesian-consistency" requirement of beliefs, the pivotal inference turns out to have no such constraint.

In order to explain the effectiveness of persuasion, we first consider the case in which all information of the voters comes from a manipulator ("monopolistic persuasion"). To invert the full information outcome, the manipulator can choose an information structure in which, roughly speaking, signals are of two possible qualities: *revealing* or *obfuscating*. When the signal is revealing, all voters observe the same signal, a in state α and b in state β. The signal is revealing with probability $1-\varepsilon$. Thus, when $\varepsilon = 0$, the election leads to the full information outcome.

However, with probability ε, the signal is obfuscating. In this case, in both states, almost all voters receive an uninformative signal z while a few voters receive an ("erroneous") signal, that is, they receive a in β and b in α. Hence, in this situation, a and b carry the opposite meaning from before.

What matters for the persuasion logic is that voters react to the closeness of the election. The closeness of the election tells voters something about the quality of the information of the others, and, in this way, also about the quality of their own signal. In the equilibrium we construct, a close election will imply that the signal of the others is of low quality (obfuscating), and, in this case, the meaning of an otherwise strong signal a in favor of α will be different and interpreted as being in favor of β, and vice versa for b.

A numerical example with 15 voters illustrates the persuasion logic. The construction uses the exact same fixed-point argument as the general analysis, showing that the same mechanism is already effective in small elections; see Section 1.4.3.

The manipulated equilibrium has some desirable properties. First, this behavior is based on a simple line of reasoning. In particular, voters will only need to interpret their own signal conditional on it being "obfuscating" and behave optimally given this interpretation (akin to so-called "sincere voting"). Second, the equilibrium is "attracting." In particular, its "basin of attraction" for the iterated best response dynamic is essentially the full set of strategy profiles: if we begin with almost any strategy profile and consider, first, the voters' best response to it and then the voters' best response to this best response, then the resulting strategy profile is arbitrarily close to the manipulated equilibrium when the number of voters is large (Proposition 2).

Further, we show that the same information structure can be used uniformly across many environments (Proposition 1). This implies that the sender does not need to know the exact details of the game. By way of contrast, existing work assumes that the manipulator knows the exact preference of each individual voter and this knowledge is indeed used. We discuss persuasion with known preferences in detail in Section 1.6.2. Finally, we show that, given the information structure, there is always one other equilibrium that yields the full-information outcome (Theorem 3).

In the second part of the paper, we consider the setting in which voters already have access to exogenous information of the form studied in Feddersen and Pesendorfer (1997). We show that, by adding information with the same signal structure as before to the exogenous information, the manipulator can still persuade the voters effectively to elect any state-contingent policy (Theorem 4). Thus, again, the additional signal structure does not need to be finely tuned to the details of the environment and is effective independent of the voters' private information. (In fact, it is shown that, when voters have exogenous private signals, then the sender needs even less information about the environment.)

In Section 2.7, we discuss the paper's contribution to the existing literature on information aggregation in elections and on voter persuasion, especially the work by Wang (2013), Alonso and Câmara (2016), Chan, Gupta, Li, and Wang (2019) and Bardhi and Guo (2018). This literature observed in particular that, with multiple receivers, the conditioning on being pivotal weakens the Bayesian consistency constraint. The main difference is that this prior work assumes that the voters' preferences and information are commonly known. Here, we allow for heterogeneous, privately known preferences and exogenous information. On the one hand, this allows capturing the canonical environment by Feddersen and Pesendorfer (1997) where, otherwise, equilibrium implies the full-information outcome. On the other hand, the persuasion mechanism here is distinct from the persuasion logic when voters' preferences are commonly known and voters can be targeted individually, as illustrated in an example in Section 1.6.2.

We note two broader implications of our analysis. First, it may be difficult for an outside observer to make a "robust" prediction. If an observer knows that voters have access to at least the information assumed in Feddersen and Pesendorfer

(1997) but cannot exclude that voters have access to additional information of the type discussed here, then no outcome can be excluded as an equilibrium prediction. Second, if one interprets an information structure with a small ε as a small departure from common knowledge, our result adds another observation to the literature on the effects of strategic uncertainty (Weinstein and Yildiz, 2007).

1.1 Model

There are $2n+1$ voters (or citizens), two policies, A and B, and two states of the world, $\omega \in \{\alpha, \beta\}$. The prior probability of α is $\Pr(\alpha) \in (0,1)$.

Voters have heterogeneous preferences. A voter's preference is described by a type $t = (t_\alpha, t_\beta) \in [-1,1]^2$, with t_ω being the utility of A in ω. The utility of B is normalized to 0; so, t_ω is the difference of the utilities from A and B in ω. The types are independently and identically distributed across voters according to a cumulative distribution function $G : [-1,1]^2 \to [0,1]$, with a strictly positive, continuous density g. The own type is the private information of the voter.

An *information structure* π is a finite set of signals S and a joint distribution of signal profiles and states that is independent of G. The conditional distribution is exchangeable with respect to the voters. In particular, there is a finite number of substates $\{\alpha_j\}_{j=1,\dots,N_\alpha}$ and $\{\beta_j\}_{j=1,\dots,N_\beta}$, such that the signals are independently and identically distributed conditional on the substates.[1] Abusing notation slightly, $\Pr(\omega_j|\omega)$ and $\Pr(s_i|\omega_j)$ denote the corresponding probabilities of the substates and the individual signal s_i, conditional on a substate. Thus, the probability of the signal profile $\boldsymbol{s} = (s_i)_{i=1,\dots,2n+1} \in S^{2n+1}$ is

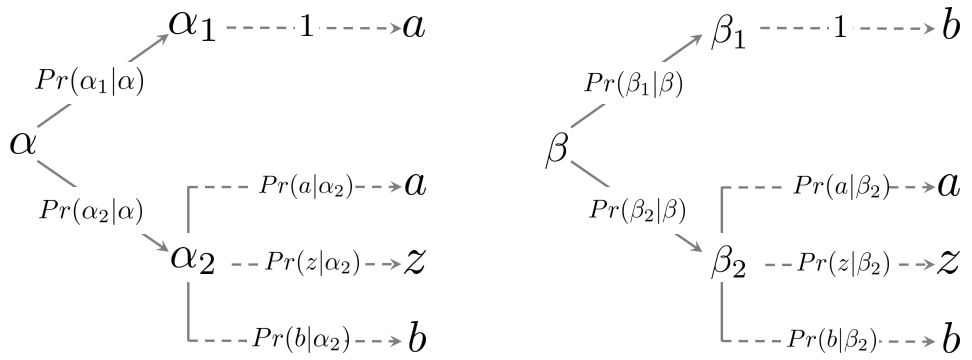
$$\Pr(\boldsymbol{s}|\omega) = \sum_j \Pr(\omega_j|\omega) \prod_{i=1,\dots,2n+1} \Pr(s_i|\omega_j). \tag{1.1}$$

The observed signal is the private information of the voter as well.

We can show our main results already with a simple class of information structures with just two substates—$\{\alpha_1, \alpha_2\}$ and $\{\beta_1, \beta_2\}$—and three conditionally independent signals in each substate—$s \in \{a, b, z\}$; this information structure is illustrated in Figure 2.1.

The voting game is as follows. First, nature draws the state, the profile of preferences types $\boldsymbol{t}$ and the profile of signals $\boldsymbol{s}$ according to G and π. Second, after observing her type and signal, each voter simultaneously submits a vote for A or B. Finally, the submitted votes are counted and the majority outcome is selected. This defines a Bayesian game.

1. The Hewitt-Savage-de Finetti theorem states that, for any exchangeable infinite sequence of random variables $(X_i)_{i=1,\dots,\infty}$ with values in some set X, there exists a random variable Y, such that the random variables X_i are independently and identically distributed conditional on Y.



Notes: Each state ω has two substates $\{\omega_1, \omega_2\}$, occurring with conditional probabilities $\Pr(\omega_j|\omega)$. Conditional on the substate ω_j, the distribution of the signals $s_i \in \{a, z, b\}$ is independent and identical with the marginal probabilities denoted by $\Pr(s|\omega_j)$ (these marginals are degenerate in α_1 and β_1) .

Figure 1.1. The main class of information structures considered in this paper

A strategy of a voter is a function $\sigma : S \times [-1,1]^2 \to [0,1]$, where $\sigma(s,t)$ is the probability that a voter of type t with signal s votes for A.

We consider only weakly undominated strategies. In particular, we require that

$$\begin{aligned} \sigma(s,t) &= 0 \quad \text{for all} \quad t = (t_\alpha, t_\beta) < (0,0), \\ \sigma(s,t) &= 1 \quad \text{for all} \quad t = (t_\alpha, t_\beta) > (0,0), \end{aligned} \tag{1.2}$$

where $t > (0,0)$ and $t < (0,0)$ are *partisans* who prefer A and B, respectively, independently of the state. Given our full support assumption on G, this rules out degenerate strategies for which either $\sigma(s,t) = 1$ for all (s,t) or $\sigma(s,t) = 0$ for all (s,t). Here, and in the following, we ignore zero measure sets when writing "*for all*".

From the viewpoint of a given voter and given any strategy σ' used by the other voters, the pivotal event piv is the event in which the realized types and signals of the other $2n$ voters are such that exactly n of them vote for A and n for B. In this event, if she votes A, the outcome is A; if she votes B, the outcome is B. In any other event, the outcome is independent of her vote. Thus, a strategy is optimal if and only if it is optimal conditional on the pivotal event.

Let $\Pr(\alpha|s, \text{piv}; \sigma')$ denote the posterior probability of α conditional on s and conditional on *being pivotal*, given the measure induced by the nondegenerate strategy σ'. The strategy σ is a best response to σ' if and only if

$$\Pr(\alpha|s, \text{piv}; \sigma') \cdot t_\alpha + (1 - \Pr(\alpha|s, \text{piv}; \sigma')) \cdot t_\beta > 0 \Rightarrow \sigma(s,t) = 1, \tag{1.3}$$

and

$$\Pr(\alpha|s,\text{piv};\sigma')\cdot t_\alpha+(1-\Pr(\alpha|s,\text{piv};\sigma'))\cdot t_\beta<0\Rightarrow\sigma(s,t)=0, \tag{1.4}$$

that is, a voter supports A if the expected value of A conditional on being pivotal is strictly positive, and supports B otherwise. Note that indifference holds only for a set of types that has zero measure. For all other types, the best response is pure. It follows that there is no loss of generality to consider pure strategies with $\sigma(s,t)\in\{0,1\}$ for all (s,t).

Thus, a symmetric, undominated, and pure Bayes-Nash equilibrium of $\Gamma(\pi)$ is a strategy $\sigma:S\times[-1,1]^2\to\{0,1\}$ that satisfies (1.2), (1.3), and (1.4), with $\sigma'=\sigma$. We refer to such a strategy simply as an *equilibrium*.

1.2 Preliminary Observations

1.2.1 Inference from the Pivotal Event

When making an inference from being pivotal, voters ask which state is more likely conditional on a tie, with exactly n voters supporting A and n supporting B. It is intuitive that a tie is evidence in favor of the substate in which the election is closer to being tied in expectation. Thus, conditional on being pivotal, a voter updates toward the substate in which the expected vote share is closer to $\frac{1}{2}$. We now verify this simple intuition and introduce some notation along the way.

For a strategy σ, the probability that a voter supports A in substate ω_j is

$$q(\omega_j;\sigma)=\sum_{s\in S}\Pr(s|\omega_j)\Pr_G(\{t:\sigma(s,t)=1\}), \tag{1.5}$$

where $q(\omega_j;\sigma)$ is the *expected vote share* of A.

Given that the signals and the types of the voters are independent conditional on the substate, the probability of a tie in the vote count is

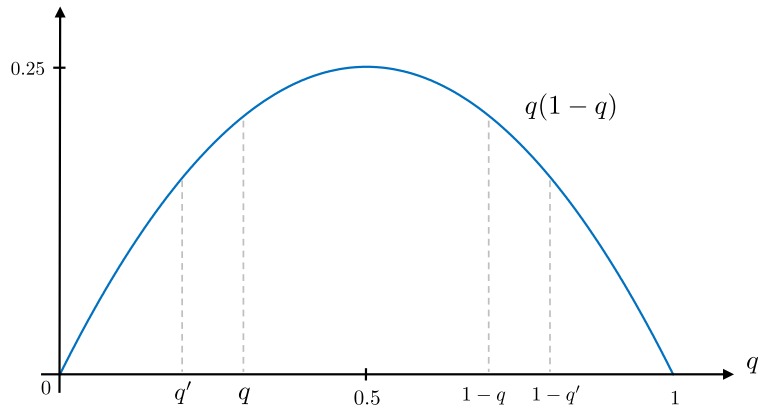
$$\Pr(\text{piv}|\omega_j;\sigma)=\binom{2n}{n}(q(\omega_j;\sigma))^n(1-q(\omega_j;\sigma))^n. \tag{1.6}$$

For any two substates ω_j and $\hat{\omega}_l$, the likelihood ratio of being pivotal is

$$\frac{\Pr(\text{piv}|\omega_j;\sigma)}{\Pr(\text{piv}|\hat{\omega}_l;\sigma)}=\left(\frac{q(\omega_j;\sigma)(1-q(\omega_j;\sigma))}{q(\hat{\omega}_l;\sigma)(1-q(\hat{\omega}_l;\sigma))}\right)^n. \tag{1.7}$$

Using the conditional independence, the posterior likelihood ratio of any two substates conditional on a signal s and the event that the voter is pivotal is

$$\frac{\Pr(\omega_j|\text{piv},s;\sigma)}{\Pr(\hat{\omega}_l|\text{piv},s;\sigma)}=\frac{\Pr(\omega_j)}{\Pr(\hat{\omega}_l)}\frac{\Pr(s|\omega_j)}{\Pr(s|\hat{\omega}_l)}\frac{\Pr(\text{piv}|\omega_j;\sigma)}{\Pr(\text{piv}|\hat{\omega}_l;\sigma)}. \tag{1.8}$$



Notes: This figure shows the function $q(1-q)$ for $q \in [0,1]$. If $|q - \frac{1}{2}| < |q' - \frac{1}{2}|$, then $q(1-q) > q(1-q')$.

Figure 1.2. The function $q(1-q)$ for $q \in [0,1]$.

We record the intuitive fact that voters update toward the substate in which the vote share is closer to 1/2, that is, in which the election is closer to being tied in expectation.

Claim 1. Take any two substates ω_j and $\hat{\omega}_l$, and any strategy σ for which $\Pr(\text{piv}|\hat{\omega}_l;\sigma) \in (0,1)$; if

$$\left| q(\omega_j;\sigma) - \frac{1}{2} \right| < \left| q(\hat{\omega}_l;\sigma) - \frac{1}{2} \right|, \tag{1.9}$$

then

$$\frac{\Pr(\text{piv}|\omega_j;\sigma)}{\Pr(\text{piv}|\hat{\omega}_l;\sigma)} > 1. \tag{1.10}$$

Proof. The function $q(1-q)$ has an inverse u-shape on $[0,1]$ and is symmetric around its peak at $q = \frac{1}{2}$, as is illustrated in Figure 1.2. So, $\left|q - \frac{1}{2}\right| < \left|q' - \frac{1}{2}\right|$ implies that $q(1-q) > q'(1-q')$. Thus, it follows from (1.7) that (1.9) implies (1.10). □

The posterior $\Pr(\alpha|s,\text{piv};\sigma)$ follows by summing over $\Pr(\alpha_j|\text{piv},s;\sigma)$.

1.2.2 Pivotal Voting

Given any strategy profile σ' used by the others, the vector of posteriors conditional on piv and s is denoted as

$$\boldsymbol{\rho}(\sigma') = (\Pr(\alpha|s,\text{piv};\sigma'))_{s\in S}. \tag{1.11}$$

This vector of posteriors is a sufficient statistic for the unique best response to σ' for all nonpartisan voter types; see (1.3) and (1.4).

Thus, given some arbitrary vector of beliefs $\boldsymbol{p} = (p_s)_{s\in S}$, let $\sigma^{\boldsymbol{p}}$ be the unique undominated strategy that is optimal if a voter with a signal s believes the probability of α to be p_s; that is, for all (s,t),

$$\sigma^{\boldsymbol{p}}(s,t) = 1 \Leftrightarrow p_s \cdot t_\alpha + (1-p_s)\cdot t_\beta > 0, \tag{1.12}$$

and (1.2) holds for the partisans.

The strategy σ is a best response to σ' if and only if $\sigma = \sigma^{\boldsymbol{p}}$ for $\boldsymbol{p} = \boldsymbol{\rho}(\sigma')$. Thus, σ^* is an equilibrium if and only if $\sigma^* = \sigma^{\boldsymbol{\rho}(\sigma^*)}$. Conversely, an equilibrium can be described by a vector of beliefs $\boldsymbol{p}^*$ that is a fixed point of $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$, that is

$$\boldsymbol{p}^* = \boldsymbol{\rho}\left(\sigma^{\boldsymbol{p}^*}\right); \tag{1.13}$$

meaning, the belief $\boldsymbol{p}^*$ corresponds to an equilibrium if, when voters behave optimally given $\boldsymbol{p}^*$ (i.e., vote according to $\sigma^{\boldsymbol{p}^*}$), the posterior conditional on being pivotal is again $\boldsymbol{p}^*$.

Equation (1.13) provides an equilibrium existence argument: the expression $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ defines a finite-dimensional mapping $[0,1]^{|S|} \to [0,1]^{|S|}$ from beliefs $\boldsymbol{p}$ into posterior beliefs $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$, and this mapping is continuous.[2] Thus, an application of Kakutani's theorem implies the existence of a fixed point $\boldsymbol{p}^*$ that solves (1.13).[3] The strategy $\sigma^{\boldsymbol{p}^*}$ is an equilibrium.[4]

The possibility of writing equilibria in terms of posteriors enables us to connect our model and results to the Bayesian persuasion literature.

1.2.3 Aggregate Preferences

A central object of the analysis is the *aggregate preference function*,

$$\Phi(p) := \Pr_G(\{t : p\cdot t_\alpha + (1-p)\cdot t_\beta > 0\}), \tag{1.14}$$

which maps a belief $p \in [0,1]$ to the probability that a random type t prefers A under p. The function Φ proves useful to express expected vote shares: if a strategy σ is optimal given beliefs $\boldsymbol{p}$—i.e., $\sigma = \sigma^{\boldsymbol{p}}$— then the expected vote share of outcome A in substate ω_j is

$$q(\omega_j;\sigma) = \sum_{s\in S} \Pr(s|\omega_j)\Phi(p_s). \tag{1.15}$$

2. To see why $\boldsymbol{\rho}\left(\sigma^{\boldsymbol{p}}\right)$ is continuous in $\boldsymbol{p}$, first, note that (1.12) implies that $\Pr_G(\{t : \sigma^{\boldsymbol{p}}(s,t) = 1\})$ is continuous in $\boldsymbol{p}$ since G has a continuous density. Second, $q(\omega_j;\sigma^{\boldsymbol{p}})$ is continuous in $\Pr_G(\{t : \sigma^{\boldsymbol{p}}(s,t) = 1\})$, given (1.5). Third, $\rho(\sigma^{\boldsymbol{p}})$ is continuous in $q(\omega_j;\sigma^{\boldsymbol{p}})$, given (1.6) and (1.8).

3. The ability to write an equilibrium as a finite-dimensional fixed point via (1.13) is a significant advantage. Similar reductions to finite dimensional equilibrium beliefs were used in related voting settings before; see Bhattacharya (2013) and Ahn and Oliveros (2012).

4. Note that, because of the partisans, $\sigma^{\boldsymbol{p}^*}$ is non-degenerate.

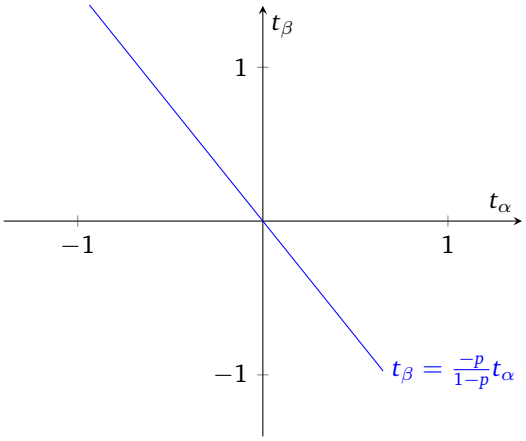


Figure 1.3. The plane of indifferent types is $t_\beta = \frac{-p}{1-p} t_\alpha$ for any given belief $p = \Pr(\alpha) \in (0,1)$.

Figure 1.3 illustrates Φ. Given p, the dashed (blue) line corresponds to the plane of indifferent types $t = (t_\alpha, t_\beta)$ with $p \cdot t_\alpha + (1-p) \cdot t_\beta = 0$. Voters having types to the north-east prefer A given p, and Φ is the measure of such types under G. The indifference plane

has a slope $-\frac{p}{1-p}$, and a change in p corresponds to a rotation of it. Given that G has a continuous density, it follows that the function Φ is continuous in p. Given that G has a strictly positive density on $[-1,1]^2$, we also have that

$$0 < \Phi(p) < 1 \quad \text{for all} \quad p \in [0,1]. \tag{1.16}$$

As observed earlier, voters having types t in the north-east quadrant prefer A for all beliefs and voters having types t in the south-west quadrant always prefer B (*partisans*). Voters having types t in the south-east quadrant prefer A in state α and B in β (*aligned voters*), and voters having types t in the north-west quadrant prefer B in state α and A in β (*contrarian voters*).

We assume throughout the paper that the distribution of types is sufficiently rich so that there is a belief p for which a majority prefers A and a belief p' for which a majority prefers B,[5] that is,

$$\Phi(p') < \frac{1}{2} < \Phi(p). \tag{1.17}$$

5. Otherwise, the analysis is trivial: if, for all beliefs $p \in [0,1]$, in expectation a majority prefers A, then, for any information structure, the vote share of A is larger than $\frac{1}{2}$ and A wins in every large election.

1.3 Large Elections: Basic Results

We consider a sequence of elections along which the electorate's size n grows. For each $2n+1$, we fix some strategy profile σ_n and calculate the probability that a policy $x \in \{A, B\}$ wins the support of the majority of the voters in state ω, denoted $\Pr\left(x|\omega; \sigma_n, n\right)$. We will be interested in the limit of $\Pr\left(x|\omega; \sigma_n^*, n\right)$, as $n \to \infty$, for equilibrium sequences $(\sigma_n^*)_{n\in\mathbb{N}}$. We first state a central observation regarding the inference from being pivotal in large elections; then, we show how this observation implies the "modern" Condorcet Jury Theorem (CJT), which we restate as a benchmark.

1.3.1 Inference in Large Elections

As a first step, we study the properties of the inference from being pivotal in a large election. We show that Claim 1 extends in an extreme form as the electorate grows large ($n \to \infty$): the event that the election is tied is infinitely more likely in the (sub-)state in which the election is closer to being tied in expectation. In fact, the likelihood ratio of the pivotal event diverges exponentially fast.

Because we want to allow the information structure to depend on n, we also include π_n in the argument. The set of substates remains fixed.

Claim 2. Consider any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$, any sequence of information structures $(\pi_n)_{n\in\mathbb{N}}$, and any two substates ω_j and $\hat{\omega}_l$ for which $\Pr\left(\text{piv}|\hat{\omega}_l; \sigma, n, \pi_n\right) \in (0,1)$ for all n. If

$$\lim_{n\to\infty} \left| q\left(\omega_j; \sigma_n, \pi_n\right) - \frac{1}{2} \right| < \lim_{n\to\infty} \left| q\left(\hat{\omega}_l; \sigma_n, \pi_n\right) - \frac{1}{2} \right|, \tag{1.18}$$

then, for any $d \geq 0$,

$$\lim_{n\to\infty} \frac{\Pr\left(\text{piv}|\omega_j; \sigma_n, \pi_n\right)}{\Pr\left(\text{piv}|\hat{\omega}_l; \sigma_n, \pi_n\right)} n^{-d} = \infty. \tag{1.19}$$

Proof. Let

$$k_n = \frac{q\left(\omega_j; \sigma_n, \pi_n\right)}{q\left(\hat{\omega}_j; \sigma_n, \pi_n\right)} \frac{\left(1 - q\left(\omega_j; \sigma_n, \pi_n\right)\right)}{\left(1 - q\left(\hat{\omega}_j; \sigma_n, \pi_n\right)\right)}.$$

From (1.7), the left-hand side of (1.19) is $\frac{(k_n)^n}{n^d}$. If (1.18) holds, then $\lim_{n\to\infty} k_n > 1$, because of the properties of $q\left(1-q\right)$ illustrated in Figure 1.2. Therefore, $\lim_{n\to\infty} \left(k_n\right)^n = \infty$. Moreover, $\left(k_n\right)^n$ diverges exponentially fast and, hence, dominates the denominator n^d, which is polynomial. □

1.3.2 Benchmark: Condorcet Jury Theorem

The model embeds a special case of the canonical voting game by Feddersen and Pesendorfer (1997) with a binary state. In the following, we restate their full-information equivalence result, assuming, at first, that signals are binary with $S = \{u, d\}$.

As in Feddersen and Pesendorfer (1997), we assume that the signals are independently and identically distributed across voters conditional on the state $\omega \in \{\alpha, \beta\}$.[6] This corresponds to the case of an information structure π^c with a single substate in each state; in the following, we identify the substate with this state. The probabilities $\Pr(s|\omega; \pi^c)$ for $s \in \{u, d\}$ and $\omega \in \{\alpha, \beta\}$ satisfy

$$1 > \Pr(u|\alpha; \pi^c) > \Pr(u|\beta; \pi^c) > 0\,; \tag{1.20}$$

that is, signal u is indicative of α, and signal d is indicative of β. We further assume that

$$\Phi(p) \text{ is strictly increasing in } p. \tag{1.21}$$

We say that the aggregate preference function is *monotone*.[7] Monotonicity (1.21) and (1.17) together imply that $\Phi(0) < \frac{1}{2} < \Phi(1)$; thus, the *full information outcome* is A in α and B in β.

Theorem 1. Feddersen and Pesendorfer (1997), Bhattacharya (2013).
Suppose that Φ is strictly increasing. Then, for every sequence of equilibria $\left(\sigma_n^*\right)_{n\in\mathbb{N}}$,

$$\begin{aligned}\lim_{n\to\infty} \Pr\left(A|\alpha; \sigma_n^*, \pi^c, n\right) &= 1,\\ \lim_{n\to\infty} \Pr\left(B|\beta; \sigma_n^*, \pi^c, n\right) &= 1.\end{aligned}$$

The proof of Theorem 1 is standard. We state it in the appendix for completeness and reference. The main observation is that the election must be equally close to being tied in both states,

$$\lim_{n\to\infty} q(\alpha; \sigma_n^*) - \frac{1}{2} = \lim_{n\to\infty} \frac{1}{2} - q(\beta; \sigma_n^*). \tag{1.22}$$

This follows in three main steps. First, voters with a signal u believe state α to be more likely than voters with a signal d. Since the probability of signal u is higher in α, this, (1.15), and the monotonicity of Φ imply a larger vote share of A in α; for all n,

$$q\left(\alpha; \sigma_n^*\right) > q\left(\beta; \sigma_n^*\right). \tag{1.23}$$

Second, in equilibrium, voters do not become certain of one of the states conditional on being tied. To see why, suppose that voters become certain the state is α,

6. Feddersen and Pesendorfer (1997) assume the existence of subpopulations and allow the signal distributions to vary across these; this is not critical. Moreover, they assume a continuum of states ω. Bhattacharya (2013) nests a binary-state version of their model. The binary state version here is a special case of the model in Bhattacharya (2013).

7. Bhattacharya (2013) says the distribution of preferences satisfies "Strong Preference Monotonicity" if (1.21) holds. He shows that monotonicity is necessary for the Condorcet Jury Theorem. If monotonicity fails, there are parameters and equilibria that do not imply the full information outcome.

that is, $\Pr(\alpha|\text{piv};\sigma_n^*) \overset{n\to\infty}{\to} 1$. Then, in both states, the vote shares would be close to $\Phi(1)$ for n sufficiently large; thus, given (1.23), for all n sufficiently large,

$$\Phi(1) > q\left(\alpha;\sigma_n^*\right) > q\left(\beta;\sigma_n^*\right) > \frac{1}{2}. \tag{1.24}$$

Equation (1.24) means that the election is closer to being tied in β. In this case, Claim 1 implies that voters update toward β conditional on being pivotal—a contradiction to the voters becoming certain of state α.

Third, since voters must not become certain of the state conditional on being pivotal, it must be that the margins of victory are equal and (1.22) holds. Otherwise, Claim 2 would imply that voters become certain of the state in which the election is closer to being tied.

Finally, (1.22) and (1.23) imply $\lim_{n\to\infty} q(\alpha;\sigma_n^*) > \frac{1}{2} > \lim_{n\to\infty} q(\beta;\sigma_n^*)$; thus, in a large election, A wins in α and B wins in β, as claimed. The proof provides the detailed argument following this outline.

Theorem 1 holds more generally for *any* sequence of information structures $(\pi_n)_{n\in\mathbb{N}}$ for which the signals are independent and identically distributed conditional on the state $\omega \in \{\alpha,\beta\}$ (i.e., there is a single substate) and for which signals do not become uninformative—that is,

$$\exists s \in S : \lim_{n\to\infty} \Pr(s|\pi_n) > 0 \text{ and } \lim_{n\to\infty} \frac{\Pr(s|\alpha;\pi_n)}{\Pr(s|\beta;\pi_n)} \neq 1. \tag{1.25}$$

Theorem 1'. Suppose Φ is strictly increasing. Then, for every sequence of information structures $(\pi_n)_{n\in\mathbb{N}}$ with a single substate and satisfying (1.25) and for every sequence of equilibria $(\sigma_n^*)_{n\in\mathbb{N}}$ given $(\pi_n)_{n\in\mathbb{N}}$,

$$\begin{aligned} \lim_{n\to\infty} \Pr\left(A|\alpha;\sigma_n^*,\pi_n,n\right) &= 1, \\ \lim_{n\to\infty} \Pr\left(B|\beta;\sigma_n^*,\pi_n,n\right) &= 1. \end{aligned}$$

1.4 Monopolistic Persuasion

We now consider the case of a sender who aims to affect the election outcome by providing information to voters, and voters have no other source of information on their own. Thus, the sender is the monopolist for information, which is the case studied in much of the literature on persuasion.

When the sender provides no information, the election outcome is trivially the outcome preferred by the majority at the prior, as determined by $\Phi(\Pr(\alpha))$. The sender can also implement the full information outcome with public signals by revealing the state. What else can the sender implement?

For example, could the sender implement a constant policy that is the opposite of what the voters prefer at the prior? Or could the sender even implement the inverse of the full information outcome? Clearly, in order to implement these policies,

the sender must provide some information to the voters, and, in fact, to implement the inverse of the full information outcome, the sender must provide sufficient information for the voters to be able to collectively distinguish the two states. On the other hand, the CJT suggests that providing information to voters may easily lead to the full information outcome, thereby suggesting that the possibility of persuasion is limited.

1.4.1 Result: Full Persuasion

Formally, we study what policies can be implemented in an equilibrium of a large election for some choice of π. This determines the set of feasible policies for a strategic sender.

The choice of the information structure π affects voters by affecting the posteriors $(\Pr(\alpha|s, \text{piv}; \sigma, \pi))_{s\in S}$. There are two effects of π. First, there is a *direct effect*; π pins down how voters learn from their signal. This effect is known from the work on persuasion. Second, there is an *indirect effect* of π because it affects the inference of the voters from being pivotal.

We show that there is no constraint on the set of feasible policies. For any state-dependent policy and for large n, there is an information structure π_n and an equilibrium σ_n for which the targeted policy wins with probability close to 1 in the respective state.[8]

Theorem 2. Take any Φ and any prior $\Pr(\alpha) \in (0,1)$: for every state-dependent policy $(x(\alpha), x(\beta)) \in \{A,B\}^2$, there exists a sequence of signal structures $(\pi_n)_{n\in\mathbb{N}}$ and equilibria $\left(\sigma_n^*\right)_{n\in\mathbb{N}}$ given $(\pi_n)_{n\in\mathbb{N}}$, such that

$$\lim_{n\to\infty} \Pr\left(x(\alpha) \,|\alpha; \sigma_n^*, \pi_n, n\right) = 1,$$
$$\lim_{n\to\infty} \Pr\left(x(\beta) \,|\beta; \sigma_n^*, \pi_n, n\right) = 1.$$

In the following, we first provide a proof for a special case of the theorem in Section 1.4.2 and illustrate it with a numerical example in Section 1.4.3. In Section 1.4.4, we discuss a general insight for persuasion in elections that underlies the result. Finally, we provide the proof for the general case in Section 1.4.5.

1.4.2 Proof: Constant Policy

This section proves Theorem 2 for the case in which Φ is monotonically increasing and the targeted policy is A in both states (i.e., Φ satisfies (1.21) and $(x(\alpha), x(\beta)) = (A,A)$). We further assume a uniform prior in order to simplify the algebra, setting $\Pr(\alpha) = \frac{1}{2}$.

8. The sender can also implement any stochastic policy by "mixing" over information structures in the appropriate manner.

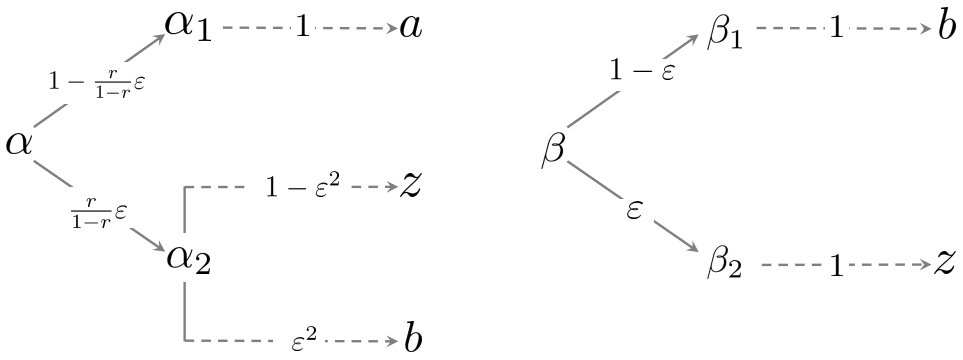


Figure 1.4. The information structure π_n^r with $\varepsilon = \frac{1}{n}$ and $r \in (0,1)$.

1.4.2.1 The Information Structure

We specialize the general information structure introduced in the model section to the one defined in Figure 1.4. Setting $\varepsilon = \frac{1}{n}$, the information structure has a single free parameter, $r \in (0,1)$, and we denote it by π_n^r.

As ε vanishes for large n, the signals are almost public in the following sense: conditional on observing any signal s, a voter believes that every other voter has received the same signal with a probability close (or equal) to 1.

Furthermore, the signals a and b reveal the state (almost) perfectly. In particular, this way the proof implies that even when constraining the sender to (almost) perfectly revealing information structures, persuasion is not constrained. In other words, the sender could be constrained to not "lie" too often.

The signal z contains only limited information since $r \in (0,1)$. When observing the signal z, a voter knows that the substate must be either α_2 or β_2. Moreover, given that a voter receives z with a probability close to 1 in either substate, we have (recall the uniform prior),

$$\lim_{n\to\infty} \Pr(\alpha|z;\pi_n^r) = \lim_{n\to\infty} \Pr(\alpha|\{\alpha_2,\beta_2\},\pi_n^r) = r. \tag{1.26}$$

1.4.2.2 Voter Inference

Clearly, for signal a,

$$\Pr(\alpha|a,\text{piv};\sigma_n,\pi_n^r) = 1. \tag{1.27}$$

Hence, in state α_1, when all voters receive a, the probability that a random citizen votes A is $\Phi(1) > \frac{1}{2}$. It follows from the weak law of large numbers that, in any equilibrium, A is elected with probability converging to 1 in state α_1.

In state β_1, all voters receive b. Conditional on the signal b alone, state β is more likely. The remaining part of this section shows that the indirect effect from the inference from being pivotal can dominate, such that there is an equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$ for which

$$\lim_{n\to\infty} \Pr(\alpha|b, \text{piv}; \sigma_n^*, \pi_n^r) = 1. \tag{1.28}$$

The proof relies on two claims. First, consider the signal z and the inference about the relative likelihood of α_2 and β_2. We show that, for *any* strategy used by the other voters, the pivotal event contains no information regarding the relative probability of α_2 and β_2 as the electorate grows large.

Claim 3. Given any $r \in (0,1)$ and any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$,

$$\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha_2; \sigma_n, \pi_n^r)}{\Pr(\text{piv}|\beta_2; \sigma_n, \pi_n^r)} = 1. \tag{1.29}$$

The proof is in the Appendix in Section 1.B.1. The pivotal event contains no information since the distribution of signals is almost identical in the two substates α_2 and β_2 (and the distribution of preference types is identical by construction). Therefore, for any strategy σ, the distribution of votes must be almost identical in the two substates; in particular, the probability of a tie is also almost the same in the two substates.[9]

Claim 3 and (1.26) imply, in particular, that for any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$,

$$\lim_{n\to\infty} \Pr(\alpha|z, \text{piv}; \sigma_n, \pi_n^r) = r. \tag{1.30}$$

Therefore, the sender can "steer" the behavior of voters with signal z by choosing r.

Next, we consider signal b and the voters' inference regarding the relative likelihood of α_2 and β_1. We show that, for this signal, the inference from the signal is dominated by the inference from being pivotal if the election is closer to being tied in state α_2 than in state β_1:

Claim 4. Take any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} |q(\sigma_n; \alpha_2, \pi_n^r) - \frac{1}{2}| < \lim_{n\to\infty} |q(\sigma_n; \beta_1, \pi_n^r) - \frac{1}{2}|; \tag{1.31}$$

then,

$$\lim_{n\to\infty} \frac{\Pr(\alpha|b, \text{piv}; \sigma_n, \pi_n^r)}{\Pr(\beta|b, \text{piv}; \sigma_n, \pi_n^r)} = \infty. \tag{1.32}$$

9. The probability that *all* voters receive signal z in state α_2 is $(1-\frac{1}{n^2})^{2n}$ and $\lim_{n\to\infty}(1-\frac{1}{n^2})^{2n} = 1$, recalling that $\lim_{n\to\infty}(1-\frac{1}{n}\frac{1}{d})^{2n} = e^{-\frac{2}{d}}$. This observation is the critical step in the proof in the appendix.

Proof. The posterior likelihood ratio is

$$\begin{aligned}\frac{\Pr(\alpha|b,\text{piv};\sigma_n,\pi_n^r)}{\Pr(\beta|b,\text{piv};\sigma_n,\pi_n^r)} &= \frac{\Pr(\alpha)}{\Pr(\beta)}\frac{\Pr\left(\alpha_2|\alpha,\pi_n^r\right)}{\Pr\left(\beta_1|\beta,\pi_n^r\right)}\frac{\Pr\left(b|\alpha_2;\pi_n^r\right)}{\Pr\left(b|\beta_1;\pi_n^r\right)}\frac{\Pr\left(\text{piv}|\alpha_2;\sigma_n,\pi_n^r\right)}{\Pr\left(\text{piv}|\beta_1;\sigma_n,\pi_n^r\right)}\\ &= \frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r\frac{1}{n}}{1-(1-r)\frac{1}{n}}\frac{\frac{1}{n^2}}{1}\frac{\Pr\left(\text{piv}|\alpha_2;\sigma_n,\pi_n^r\right)}{\Pr\left(\text{piv}|\beta_1;\sigma_n,\pi_n^r\right)}\\ &\approx \frac{\Pr\left(\text{piv}|\alpha_2;\sigma_n,\pi_n^r\right)}{\Pr\left(\text{piv}|\beta_1;\sigma_n,\pi_n^r\right)}n^{-3}. \qquad (1.33)\end{aligned}$$

For the approximation on the last line we used that the prior is uniform. Given (1.31), equation (1.32) follows from applying Claim 2 for $d = 3$. □

Thus, for any sequence of strategies that satisfies (1.31), the critical posterior with signal b satisfies the desired property (1.28).

1.4.2.3 Fixed Point Argument

By the richness assumption on Φ (see (1.17)), there is some $\hat{r}$ such that $\Phi(\hat{r}) = \frac{1}{2}$. We will show that, for the information structure $\pi_n^{\hat{r}}$ and n large enough, there is an equilibrium in which A receives a strict majority of votes in both states in expectation.

The basic idea is this: the choice of $\hat{r}$ and (1.30) imply that the vote shares in states α_2 and β_2 are close to $\Phi(\hat{r}) = \frac{1}{2}$. Moreover, in equilibrium, it will be the case that A receives a strict majority of votes in state β_1. Hence, the election is closer to being tied in α_2 than in β_1. Therefore, by Claim 4, voters with signal b become convinced that the state is α; thus, the vote share of A in β_1 is close to $\Phi(1) > \frac{1}{2}$.

Recall that equilibrium is equivalently characterized by a vector of beliefs, $\boldsymbol{p}^* = (p_a^*, p_z^*, p_b^*)$, such that $\boldsymbol{p}^* = \boldsymbol{\rho}\left(\sigma^{\boldsymbol{p}^*}\right)$; see (1.13). Now, for any $\delta > 0$, let

$$\mathrm{B}_\delta = \left\{\boldsymbol{p} \in [0,1]^3 \mid |\boldsymbol{p} - (1,\hat{r},1)| \leq \delta\right\},$$

so that B_δ is the set of beliefs at most δ away from $(1,\hat{r},1)$. Take any $\boldsymbol{p} \in \mathrm{B}_\delta$ and the corresponding strategy $\sigma^{\boldsymbol{p}}$. Since $\Phi(1) > \frac{1}{2}$, this means that A receives a strict majority of votes in the states α_1 and β_1 for δ small enough. In the states α_2 and β_2, (almost) all voters observe signal z, so $q(\alpha_2;\sigma^{\boldsymbol{p}},\pi_n^{\hat{r}}) \approx \Phi(\hat{r})$ and $q(\beta_2;\sigma^{\boldsymbol{p}},\pi_n^{\hat{r}}) \approx \Phi(\hat{r})$. Since $\Phi(\hat{r}) = \frac{1}{2}$, the vote share for A is approximately $\frac{1}{2}$.

Now, we show that our two previous claims, Claim 3 and 4, imply that—given $\sigma^{\boldsymbol{p}}$—the posterior conditional on being pivotal is again in B_δ, for any $\boldsymbol{p} \in \mathrm{B}_\delta$, any sufficiently small δ, and any sufficiently large n:

Claim 5. For any δ sufficiently small, there exists $n(\delta)$ s.t., for all $n \geq n(\delta)$,

$$\forall \boldsymbol{p} \in B_\delta : \boldsymbol{\rho}\left(\sigma^{\boldsymbol{p}};\pi_n^{\hat{r}},n\right) \in B_\delta. \qquad (1.34)$$

Proof. Take any $\boldsymbol{p} \in \mathrm{B}_\delta$ and its corresponding behavior $\sigma^{\boldsymbol{p}}$. For the posterior following signal a it is immediate that, for all δ and n,

$$\boldsymbol{\rho}_a\left(\sigma^{\boldsymbol{p}}; \pi_n^{\hat{r}}, n\right) = 1; \tag{1.35}$$

see (1.27). Secondly,

$$\lim_{n\to\infty} \boldsymbol{\rho}_z\left(\sigma^{\boldsymbol{p}}; \pi_n^{\hat{r}}, n\right) = \hat{r}, \tag{1.36}$$

follows from Claim 3 for all δ; see (1.30).

Finally, for δ small enough and n large enough, the election is closer to being tied in α_2 than in β_1,

$$\forall \boldsymbol{p} \in \mathrm{B}_\delta\colon\ |q(\alpha_2; \sigma^{\boldsymbol{p}}, \pi_n^{\hat{r}}) - \frac{1}{2}| < |q(\beta_1; \sigma^{\boldsymbol{p}}, \pi_n^{\hat{r}}) - \frac{1}{2}|. \tag{1.37}$$

To see why, note that for n large enough, $q(\alpha_2; \sigma^{\boldsymbol{p}}, \pi_n^{\hat{r}}) \approx \Phi\left(p_z\right)$ and $q(\beta_1; \sigma^{\boldsymbol{p}}, \pi_n^{\hat{r}}) = \Phi\left(p_b\right)$ since almost all voters receive z in α_2 and all voters receive b in β_1. In addition, by the continuity of Φ, for δ small enough, we have that $\Phi\left(p_z\right) \approx \Phi\left(\hat{r}\right)$ and $\Phi\left(p_b\right) \approx \Phi\left(1\right)$. Finally, (1.37) follows then from $\Phi\left(\hat{r}\right) = \frac{1}{2}$ and $\Phi\left(1\right) > \frac{1}{2}$.

Now, it follows from (1.37) and from Claim 4 that

$$\lim_{n\to\infty} \boldsymbol{\rho}_b\left(\sigma^{\boldsymbol{p}}; \pi_n^{\hat{r}}, n\right) = 1. \tag{1.38}$$

Thus, the claim follows from (1.35),(1.36), and (1.38). □

Since $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ is continuous in $\boldsymbol{p}$ by the arguments after (1.13), it follows from (1.34) and Kakutani's theorem that there exists a fixed point $\boldsymbol{p}_n^* \in B_\delta$ for all n large enough. By the arguments from the proof of Claim 5,

$$\lim_{n\to\infty} \boldsymbol{p}_n^* = \left(1, \hat{r}, 1\right), \tag{1.39}$$

see (1.35), (1.36), and (1.38). Finally, for the corresponding sequence of equilibrium strategies, $(\sigma^{\boldsymbol{p}_n^*})_{n\in\mathbb{N}}$, the policy A wins in both states; this follows from (1.39), which implies that voters with signals a and b are supporting A with a probability converging to $\Phi\left(1\right) > \frac{1}{2}$, and from the weak law of large numbers.

This completes the proof of the theorem for the special case in which Φ is monotone, the targeted policy is A in both states, and the prior is uniform. When the prior is not uniform, the only piece of the argument that needs to be adjusted is the choice of r. For a general prior $\Pr\left(\alpha\right) \neq \frac{1}{2}$, the value of r should be such that

$$\frac{\Pr\left(\alpha\right) r}{\Pr\left(\alpha\right) r + \left(1 - \Pr\left(\alpha\right)\right)\left(1 - r\right)} = \hat{r}, \tag{1.40}$$

with $\Phi(\hat{r}) = \frac{1}{2}$.

1.4.3 Numerical Example with 15 voters

Let $\Phi(p) = p$ for all $p \in [0,1]$.[10] Further, we set $\Pr(\alpha) = \frac{1}{2}$ and let the information structure be π_n^r with $r = \frac{1}{2}$. In the Appendix in Section 1.B.2, we show that under these primitives, when there are at least $2n+1 = 15$ voters, there is an equilibrium σ_n^* for which A is elected with a probability larger than 99.9% in the states α_1 and β_1. Therefore, the overall probability of A being elected exceeds $0.999(1 - \frac{1}{n})$, which is larger than 85% when there are at least $2n+1 = 15$ voters.

To do so, we show that under the specified primitives, when $n \geq 7$, the best response induces a self-map $\boldsymbol{\rho}$ on the set of beliefs $\boldsymbol{p} = (p_a, p_z, p_b) \in [0,1]^3$ for which $p_a \geq 0.95$, $p_z \in [0.32, 0.68]$, and $p_b \geq 0.95$. Then, an application of Kakutani's theorem yields an equilibrium in which voters with an a-or b-signal vote A with a probability of at least 95%. Evaluation of the binomial distribution $\mathrm{B}(2n+1, x)$ for $x \geq 0.95$ shows that indeed A receives a majority of the votes with probability larger than 99.9% in the states α_1 and β_1 where all voters receive signal a or b.

1.4.4 Persuasion in Elections

As noted, voters' behavior is determined by their critical belief, $\Pr(\alpha|s, \text{piv}; \sigma, \pi)$, implying a close connection to the standard information design and persuasion model. The signal structure π affects voters' belief directly via the inference from s and indirectly via the inference from being pivotal. Bayesian consistency is understood to constrain a sender's ability to affect the signal inference by choice of π; however, the indirect effect is much less constrained.

Bayesian consistency—or the law of iterated expectation—requires that

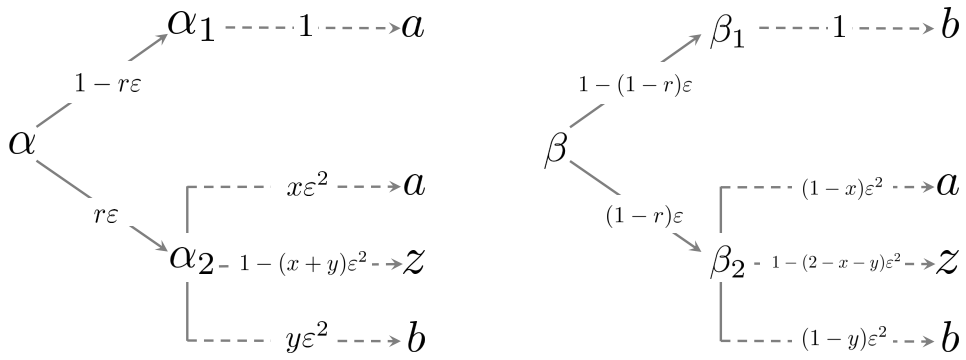
$$\Pr(\alpha) = \sum_{s \in S} \left[\Pr(s, \text{piv}) \Pr(\alpha|s, \text{piv}) + \Pr(s, \neg\text{piv}) \Pr(\alpha|s, \neg\text{piv}) \right], \tag{1.41}$$

where $\Pr(\alpha|s, \neg\text{piv}; \sigma, \pi)$ is the posterior conditional on not being pivotal, and we omitted (σ, π). With a single voter, $\Pr(\text{piv}) = 1$, and so the expected critical belief is constraint to be the prior. However, with many voters, $\Pr(\text{piv})$ becomes small, and, consequently, (1.41) imposes only a small constraint.

The effectiveness of "pivotal persuasion" has been observed before in a setting with known preferences and no private information by the voters; see our discussion of the related literature in Section 1.6.2; especially Chan et al. (2019) and Bardhi and Guo (2018).

Intuitively, what matters is that voters react to the closeness of the election. The closeness of the election tells voters something about the information of others, and,

10. We provide an explicit example of a preference distribution G that induces $\Phi(p) = p$ for all p. Since, therefore, $\Pr(t : t_\alpha > 0, t_\beta < 0) = 1$, the example fails the assumption that G has a strictly positive density on $[-1,1]^2$. This simplifies the presentation and one can find a nearby example with full support.



Notes: The parameter r controls the posterior after z and the parameters x and y control the beliefs after a and b, respectively, conditional on being in substate α_2 or β_2.

Figure 1.5. The information structure $\pi_n^{x,r,y}$ with $\varepsilon = \frac{1}{n}$ and $(x,r,y) \in [0,1]^3$.

in this way, about the quality of the signal structure. The quality of the signal structure, in turn, affects the meaning of the own information.

In our construction, one may interpret the signal structure π^r as releasing either a high quality signal—in substates $\{\alpha_1, \beta_1\}$—or a low quality signal—in substates $\{\alpha_2, \beta_2\}$. The closeness of the election depends on the signal quality. In particular, when the quality of the signal structure is high, all voters observe the same revealing signal and the election is far from being close. Conversely, when the election is close, this is because the quality of the signal is low. In this case, most voters learn that the signal quality is low but some may receive erroneous messages. In particular, when the election is close and the signal quality low, the meaning of a b signal changes from being indicative of β to being an erroneous signal that is indicative of α.

The pivotal voting model considers the extreme case where voters react perfectly to the closeness of the election; it illustrates the extreme effectiveness of persuasion in this case. One may conjecture that, in a setting in which voters react less sensitively, persuasion is still effective but, presumably, less extreme.

1.4.5 Sketch of Proof: General Policy

Now, we allow for non-monotone Φ and show that the sender can implement any intended state-dependent policy, including the one that inverts the full-information outcome.

For this, we consider the information structure depicted in Figure 1.5. The signals are (almost) public, similar to the information structure in the previous section from Figure 1.4. Moreover, as before, the signals a and b reveal the state (almost) perfectly. The signal z contains only limited information since $r \in (0,1)$. When ob-

serving the signal z, a voter knows that the substate must be either α_2 or β_2, and her belief conditional on signal z is given by

$$\lim_{n\to\infty} \frac{\Pr(\alpha|z;\pi_n^{x,r,y})}{\Pr(\beta|z;\pi_n^{x,r,y})} = \lim_{n\to\infty} \frac{\Pr(\alpha|\{\alpha_2,\beta_2\};\pi_n^{x,r,y})}{\Pr(\beta|\{\alpha_2,\beta_2\};\pi_n^{x,r,y})} = \frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r}. \quad (1.42)$$

We prove Theorem 2 by showing that by choosing the parameters $(x,r,y) \in [0,1]^3$ appropriately, the sender can implement almost any belief μ_α in state α and any belief μ_β in state β as $n \to \infty$, in the sense that, with probability close to one, almost all voters will have such beliefs conditional on being pivotal.

Lemma 1. Let $\hat{r}$ solve $\Phi(\hat{r}) = \frac{1}{2}$ and suppose $\hat{r} \notin \{0,1\}$. Take any $(\mu_\alpha,\mu_\beta) \in [0,1]^2$ with $\Phi(\mu_\alpha) \neq \frac{1}{2}$ and $\Phi(\mu_\beta) \neq \frac{1}{2}$ and choose $(x,r,y) \in [0,1]^3$ as the solutions to[11]

$$\frac{\hat{r}}{1-\hat{r}}\frac{x}{1-x} = \frac{\mu_\alpha}{1-\mu_\alpha}, \quad (1.43)$$

$$\frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r} = \frac{\hat{r}}{1-\hat{r}}, \quad (1.44)$$

$$\frac{\hat{r}}{1-\hat{r}}\frac{y}{1-y} = \frac{\mu_\beta}{1-\mu_\beta}. \quad (1.45)$$

Then, there exists a sequence of equilibria $(\sigma_n^*)_{n\in\mathbb{N}}$ given $(\pi_n)_{n\in\mathbb{N}} = (\pi_n^{x,r,y})_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv},a;\sigma_n^*,\pi_n\right) = \mu_\alpha, \quad (1.46)$$

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv},z;\sigma_n^*,\pi_n\right) = \hat{r}, \quad (1.47)$$

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv},b;\sigma_n^*,\pi_n\right) = \mu_\beta. \quad (1.48)$$

The lemma is proven in the Appendix in Section 1.B.3, using ideas similar to those used earlier. First, as before, voters with signals z do not update conditional on being pivotal as $n \to \infty$ in any equilibrium, and r is then chosen such that, in substates α_2 and β_2, the vote share of A is close to $\frac{1}{2}$ in every equilibrium. Second, we show that there are equilibria in which voters with signals a and b behave according to the beliefs μ_α and μ_β. By the choice of the beliefs, with this behavior, there is either a strict majority for A or B in the substates α_1 and β_1; thus, the election is closer to being tied in α_2 and β_2 than in α_1 and β_1. Thus, conditional on being pivotal, voters with signals a and b believe that they are in substates α_2 and β_2, and, interpreting their signals conditional on these substates, their critical posteriors are as given in the lemma.

The lemma implies Theorem 2: the richness assumption (1.17) states that there is a belief p for which a majority prefers A in expectation and a belief p' for which

11. For $\mu_\alpha = 1$, let $x = 1$, and for $\mu_\beta = 1$, let $y = 1$ such that the following equations hold in the extended reals, using the convention that $\frac{1}{0} = \infty$.

a majority prefers B in expectation—that is, $\Phi(p) > \frac{1}{2} > \Phi(p')$. Thus, given belief p', it follows from the weak law of large numbers that B is elected with probability converging to 1. Given belief p, it follows from the weak law of large numbers that A is elected with probability converging to 1. Hence, the sender can implement any state-contingent policy $(x_\alpha, x_\beta) \in \{A, B\}^2$ by implementing belief p in any state ω for which $x_\omega = A$ and by implementing belief p' in any state for which $x_\omega = B$.

1.4.6 Robustness

In this section, we discuss the robustness of the persuasion result in Theorem 2. In particular, we ask: can the sender persuade the voters even when he does not know the exact details of the environment? How "stable" is the equilibrium? Are there other equilibria?

1.4.6.1 Robustness: Detail-Freeness

In this section, we show that in order to persuade the voters, the signal structure does not need to be finely tuned to the details of the environment. Suppose that the prior and the preference distribution are such that

$$|\Phi(0) - \frac{1}{2}| > |\Phi(\Pr(\alpha)) - \frac{1}{2}|, \tag{1.49}$$

$$|\Phi(1) - \frac{1}{2}| > |\Phi(\Pr(\alpha)) - \frac{1}{2}|; \tag{1.50}$$

therefore, when the citizens vote optimally given their beliefs, the election is closer to being tied when they they are uninformed and hold the prior belief relative to when know the state.

Proposition 1. Take $r = 1$ and $(x, y) \in \{0, 1\}^2$. For any prior and preference distribution satisfying (1.49) and (1.50), there is a sequence of equilibria $\left(\sigma_n^*\right)_{n\in\mathbb{N}}$ given the sequence of signal structures $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv}, a; \sigma_n^*\right) = x, \tag{1.51}$$

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv}, z; \sigma_n^*\right) = \Pr(\alpha), \tag{1.52}$$

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv}, b; \sigma_n^*\right) = y. \tag{1.53}$$

The proposition implies that the sender can implement any policy using a single signal structure that works uniformly across the large set of priors and preference distributions satisfying (1.49) and (1.50). For example, the constant policy A is implemented by choosing $x = y = 1$, which leads to an equilibrium in which A has a vote share $\Phi(1)$ as the election becomes large.

The proof is in the appendix in Section 1.B.4. The basic idea is that, given this signal, the vote shares are close to $\Phi(\Pr(\alpha))$ in states α_2 and β_2. Hence, by assumptions (1.49) and (1.50), if voters behave according to the posteriors x and y in states

α_1 and β_1, the election is closer to being tied in α_2 and β_2 than in α_1 and β_1. Thus, just as before, conditional on being pivotal, voters with signals a and b believe that they are in states α_2 and β_2, and—interpreting their signals conditional on these substates—their critical posteriors are as given in the proposition.

A similar argument implies that the signal structure from Lemma 1 is also effective when the actual environment is slightly different: When the prior and Φ is slightly different from the one used to calculate (x, r, y), then there is still an equilibrium close-by with critical beliefs that are close to μ_α, $\hat{r}$, and μ_β, provided that vote shares at the critical beliefs imply that the election is still closer to being tied in states α_2 and β_2 than in states α_1 and β_1.

Random Signal Quality. Note that the signal from Proposition 1 matches the description in the introduction. In particular, we can swap the timining in the description of the signal. Rather than choosing the "quality" of the signal after the state of nature has realized, one can first choose randomly whether the signal is "revealing" or "obfuscating" and then, if it is revealing, send a signal corresponding to the realized state of nature to all voters (as in substates α_1 and β_1), and, if it is obfuscating, send the signals z or b in α and z or a in β (as in substates α_2 and β_2 when $x = 0$ and $y = 1$).

1.4.6.2 Robustness: Basin of Attraction

We show that, for a large set of initial strategies, an iterated best response leads quickly to the "manipulated equilibrium" of Theorem 2 described earlier.

Let (μ_α, μ_β) be any pair of beliefs with $\Phi(\mu_\alpha) \neq \frac{1}{2}$ and $\Phi(\mu_\beta) \neq \frac{1}{2}$. By Lemma 1, there is a sequence of information structures $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ and equilibria $(\sigma_n^*)_{n\in\mathbb{N}}$ that implements the pair of beliefs as $n \to \infty$, in the sense that, with probability close to 1, almost all voters will have such beliefs conditional on being pivotal. Hence, by choosing (μ_α, μ_β) appropriately, a sender can implement any desired policy. The next result shows that, for almost any strategy σ, the twice iterated best response is arbitrarily close to σ_n^* when n is large, in the sense that the posteriors conditional on being tied are close to (μ_α, μ_β).

First, let us define the twice iterated best response: take any belief $\boldsymbol{p}$ and the strategy $\sigma^{\boldsymbol{p}}$ that is optimal given these beliefs. Then, $\sigma^{\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})}$ is the best response to $\sigma^{\boldsymbol{p}}$ and is optimal given the beliefs

$$\boldsymbol{\rho}^1(\boldsymbol{p}) = \boldsymbol{\rho}(\sigma^{\boldsymbol{p}}), \tag{1.54}$$

where $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ is the vector of the posteriors conditional on the pivotal event and the signals. In the same way, $\sigma^{\boldsymbol{\rho}(\sigma^{\boldsymbol{\rho}^1(\boldsymbol{p})})}$ is the best response to $\sigma^{\boldsymbol{\rho}^1(\boldsymbol{p})}$ (so it is the twice iterated best response to $\sigma^{\boldsymbol{p}}$) and is optimal given the beliefs

$$\boldsymbol{\rho}^2(\boldsymbol{p}) = \boldsymbol{\rho}(\sigma^{\boldsymbol{\rho}^1(\boldsymbol{p})}). \tag{1.55}$$

Proposition 2 shows that for almost any $\boldsymbol{p}$, we have $|\boldsymbol{\rho}^2(\boldsymbol{p}) - (\mu_\alpha, \hat{r}, \mu_\beta)| < \epsilon$ when n is sufficiently large. This means that the twice iterated best response is arbitrarily close to the manipulated equilibrium σ_n^* since the equilibrium is consistent with the belief $\boldsymbol{\rho}(\sigma_n^*) \approx (\mu_\alpha, \hat{r}, \mu_\beta)$; see (1.13).

Proposition 2. Take any beliefs $(\mu_\alpha, \mu_\beta) \in [0,1]^2$ with $\Phi(\mu_\alpha) \neq \frac{1}{2}$ and $\Phi(\mu_\beta) \neq \frac{1}{2}$ and the corresponding information structures $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ from Lemma 1.

For any $\delta > 0$, there is some $B \subset [0,1]^3$ with Lebesgue-measure of at least $1-\delta$ and some $\bar{n} \in \mathbb{N}$ such that, for all $n \geq \bar{n}$,

$$\forall \boldsymbol{p} \in B : |\boldsymbol{\rho}^2(\boldsymbol{p}) - (\mu_\alpha, \hat{r}, \mu_\beta)| < \delta. \tag{1.56}$$

The proof is in Section 1.B.5 in the Appendix. The proof also implies that, for "almost any" strategy σ—even those that are not optimal given some belief $\boldsymbol{p}$—the twice iterated best reply is arbitrarily close to the manipulated equilibrium σ_n^* when n is large, where the genericity requirement is with respect to the induced vote shares; see condition (1.102), replacing $\sigma^{\boldsymbol{p}}$ by σ.

Simple Reasoning. Proposition 2 illustrates that a simple reasoning underlies the manipulated equilibrium σ_n^*. The result loosely relates to the concepts of level k-thinking and level-k-implementability (De Clippel, Saran, and Serrano (2019)). The theorem implies that, for almost any strategy (a "behavioral anchor"), the strategies that are consistent with level-2-thinking are close to the manipulated equilibrium. In this sense, any state-dependent target policy $(x(\alpha), x(\beta)) \in \{A,B\}^2$ is level-2-implementable.[12]

1.4.6.3 Other Equilibria

Proposition 2 shows that the basin of attraction of the iterated best response of an arbitrarily small neighborhood of the manipulated equilibria consists of almost all strategies when n is large enough. However, this still leaves open the possibility that there are other equilibria, such that if we begin exactly at such a strategy profile, the best response dynamic stays there. In the working paper version, Heese and Lauermann (2019, Theorem 4),[13] we show that this is indeed the case. There exists another equilibrium and that equilibrium is not "manipulated" but implements the full information outcome as $n \to \infty$. We restate the result here:

12. De Clippel, Saran, and Serrano (2019) consider a different notion of level-2-implementability that demand that there is *some* behavioral anchor such that *any* profile of strategies that are level-1-consistent or level-2-consistent for this anchor implement a given social choice function. Here, almost any strategy can be such an anchor.

13. The working paper is publicly available here https://ideas.repec.org/p/bon/boncrc/crctr224_2019_128.html.

Theorem 3. Let Φ be strictly increasing. For all information structures $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ with $(x,r,y) \in (0,1)^3$, there exists an equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$ for which the full information outcome is elected as $n \to \infty$,

$$\lim_{n\to\infty} \Pr(A|\alpha; \sigma_n^*, \pi_n, n) = 1,$$
$$\lim_{n\to\infty} \Pr(B|\beta; \sigma_n^*, \pi_n, n) = 1.$$

Intuition. Note that the signal π_n almost always sends an (almost) perfectly revealing signal when n is large. Hence, there is a sequence of strategies (e.g. given by sincere voting) for which the full-information outcome is elected as $n \to \infty$. The question is if such a sequence of strategies can be an equilibrium sequence. The theorem shows that, whenever Φ is monotone, the answer is yes. This is easy to see in the extreme case when voters have a common type t, and, hence, have common interests. A result of McLennan (1998) shows that, with common interest, the utility maximizing symmetry strategy is a symmetric equilibrium. Hence, for this case, the existence of a sequence of strategies that yields the full-information outcome immediately implies the existence of an equilibrium sequence that yields it as well.

1.5 Persuasion of Privately Informed Voters

Recall the binary information structure from the Condorcet Jury Theorem, defined by the signal probabilities $\Pr(s|\omega)_{\omega\in\{\alpha,\beta\}}$ for $s \in \{u,d\}$ such that (1.20) holds. We will think of this as exogenous private information that is held by the voters and denote this information structure by π^c. We say that an information structure π with signal set S is an *independent expansion* of π^c if it is the product of π^c and some additional signal structure π^p that is exchangeable, as before.[14]

We think of the expansion as resulting from additional information π^p that is provided by a sender to voters who also receive private signals from π^c. By considering only independent expansions, we do not allow the sender's signal to condition directly on the realization of π^c. As before, we also do not allow the sender to elicit the voters' private information (the preference type and the signal). We assume that the preferences of the voters are such that the aggregate preference function Φ is strictly increasing so that the CJT holds (Theorem 1) and, without an additional signal, the unique equilibrium outcome is the full information outcome as the electorate grows large.

14. More formally, π is an independent expansion if there exists an information structure π^p with signal set S_2 and substates $\{\alpha_1,\dots,\alpha_{N_\alpha}\}$ and $\{\beta_1,\dots,\beta_{N_\beta}\}$ such that $S = \{u,d\} \times S_2$ and

$$\Pr(\boldsymbol{s}|\omega_j;\pi) = \Pr(\boldsymbol{s}_1|\omega;\pi^c)\Pr(\boldsymbol{s}_2|\omega_j;\pi^p) \tag{1.57}$$

for all $\omega_j \in \{\alpha_1,\dots,\alpha_{N_\alpha}\} \cup \{\beta_1,\dots,\beta_{N_\beta}\}$ and all $\boldsymbol{s} = (\boldsymbol{s}_1,\boldsymbol{s}_2) \in (\{u,d\} \times S_2)^{2n+1}$.

What outcomes can the sender implement when the voters have exogenous signals and how should he communicate with the voters? Clearly, to implement any policy other than the full information outcome, the sender has to communicate with the voters in some way. Consider a sender who communicates with *public signals* $s_2 \in S_2$, meaning, that the signals are commonly received by all the voters.[15] When the voters receive a public signal s_2, this shifts the common belief from the prior $\Pr(\alpha)$ to $\Pr(\alpha|s_2)$. Since the CJT holds for any common prior, in the subgame following any public signal, the full information outcome is elected with probability converging to 1, as $n \to \infty$.[16] So, in order to implement any outcome other than the full information outcome, the sender has to communicate privately with the voters.

1.5.1 Result: Full Persuasion

The following theorem shows that there exists an independent expansion of the private information of the voters that allows implementing any state-dependent policy—even the policy that inverts the full-information outcome.

Theorem 4. Take any exogenous private signals π^c of the voters satisfying (1.20) and any strictly increasing Φ. For every state-dependent policy $(x(\alpha), x(\beta)) \in \{A, B\}^2$, there exists a sequence of independent expansions $(\pi_n)_{n\in\mathbb{N}}$ of π^c and equilibria $\left(\sigma_n^*\right)_{n\in\mathbb{N}}$ given $(\pi_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \Pr\left(x(\alpha) \,|\alpha; \sigma_n^*, \pi_n, n\right) = 1,$$
$$\lim_{n\to\infty} \Pr\left(x(\beta) \,|\beta; \sigma_n^*, \pi_n, n\right) = 1.$$

The next two sections provide an extensive sketch of the arguments establishing the theorem. In particular, the original signals from the previous section are sufficient, namely, π_n^r, as in Figure 1.4 can be chosen as an additional signal to implement equilibria in which A wins in both states, and $\pi_n^{x,r,y}$ from Figure 1.5 with $x = 0$ and $y = 1$ can be chosen to implement a policy that inverts the full-information outcome. Thus, the sender does not need to know whether agents have private information nor how much private information they have—the same signal structure works uniformly across environments.

1.5.2 Sketch of the Proof: Constant Policy

We show that the same signal structure π_n^r from Figure 1.4 leads to an equilibrium in which A wins in both states—even when voters have private signals.

15. Alonso and Câmara (2016) have studied persuasion with public signals when voters do not have exogenous private signals.

16. To be precise, the CJT only applies to any non-degenerate prior $\Pr(\alpha) \in (0, 1)$. However, if the sender reveals the state publicly, such that $\Pr(\alpha|s) \in \{0, 1\}$, trivially, the full-information outcome is elected as $n \to \infty$.

The critical observation in the proof is that the vote shares in α_2 and β_2 are uniquely determined across all equilibria and parameter by an equal-margin-of-victory condition.

Claim 6. Let Φ be strictly increasing. Suppose that the additional information is given by π_n^r, as in Figure 1.4. Then, there is some M with

$$0 < M < \Phi(1) - \frac{1}{2} \tag{1.58}$$

such that, for every $r \in (0,1)$ and every equilibrium sequence (σ_n^*) given π_n^r,

$$\lim_{n\to\infty} q(\sigma_n^*; \alpha_2, \pi_n^r) - \frac{1}{2} = \lim_{n\to\infty} \frac{1}{2} - q(\sigma_n^*; \beta_2, \pi_n^r) = M. \tag{1.59}$$

For the proof, see Section 1.C.2 in the Appendix. The idea is the following: given π_n^r, in substates α_2 and β_2, every voter receives the additional signal z with probability converging to 1. Voters who received z know that either α_2 or β_2 holds and that almost all other voters got a signal z as well. Hence, from their perspective, it is close to common knowledge that the game is close to a game with a binary state and binary signals π^c, as in the original setting of the CJT. Now, the proof of the CJT showed that the election must be equally close to being tied in expectation—see (1.22)—and the same arguments implies (1.59) here.

Now, one can show that there is a sequence of equilibria in which the vote share of A in state β_1 approaches its maximum, $\Phi(1)$, and thus

$$\lim_{n\to\infty} q(\sigma_n^*; \beta_1, \pi_n) - \frac{1}{2} = \Phi(1) - \frac{1}{2}. \tag{1.60}$$

Comparing (1.59) and (1.60), in this equilibrium sequence, the election is closer to being tied in α_2 than in β_1. Hence, it follows from Claim 2 that

$$\lim_{n\to\infty} \frac{\Pr\left(\text{piv}|\beta_1; \sigma_n^*\right)}{\Pr\left(\text{piv}|\alpha_2; \sigma_n^*\right)} = 0. \tag{1.61}$$

Moreover, it follows also from Claim 2 that the inference from the pivotal event dominates the direct inference from the signal;[17] so, a voter with additional signal $s_2 = b$ becomes convinced that the state is α_2 for either realization of the private signal $s_1 \in \{u, d\}$,

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv}, s_1, s_2 = b; \sigma_n^*\right) = 1. \tag{1.62}$$

Since all voters observe the additional signal $s_2 = b$ in state β_1, it follows that the vote share converges to $\Phi(1)$, as claimed in (1.60). Finally, it is clear that such an equilibrium sequence leads to outcome A in both states with probability converging to 1.

17. See the proof of the analogous Claim 4.

Note that the basic idea here is similar to the one in Section 1.4.2 without the private signal. Here, Claim 6 pins down behavior in states α_2 and β_2, analogously to the implication of the previous Claim 3. Then, there is an equilibrium in which A receives a strict majority in β_1. In both settings, the equilibrium is supported by the fact that the election is closer to being tied in α_2 than in β_1, so that, conditional on being pivotal, voters with signal b become convinced that the state is α_2.

1.5.3 Sketch of Proof: General Policy

The signal $\pi_n^{x,r,y}$ from Figure 1.5 can again be used to implement any intended policy by the appropriate choice of $(x,r,y) \in [0,1]^3$. The proof of this general result utilizes a lemma analogous to the previous Lemma 1, stated as Lemma 3 in the appendix.

In particular, as before, the policy that inverts the full-information outcome can be implemented by choosing the additional signal with $x = 0$, $y = 1$, and any *arbitrary* $r \in (0,1)$: for any such choice, we show that there is a sequence of equilibria (σ_n^*) in which the posterior probabilities conditional on being pivotal and the additional signals a and b are close to 0 and 1, respectively. Moreover, since the private signals are boundedly informative, it follows that, for $s_1 \in \{a,b\}$,

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv}, s_1, s_2 = a; \sigma_n^*\right) = 0, \tag{1.63}$$

$$\lim_{n\to\infty} \Pr\left(\alpha|\text{piv}, s_1, s_2 = b; \sigma_n^*\right) = 1. \tag{1.64}$$

Thus, since all voters observe signals a and b in the substates α_1 and β_1, respectively, the equilibrium vote shares converge to $\Phi(0) < 1/2$ and $\Phi(1) > 1/2$, with the inequalities from Φ satisfying (1.17). Therefore, the weak law of large numbers implies that B wins in state α_1 and A wins in state β_1, thereby establishing the existence of an equilibrium that inverts the full-information outcome.

1.5.4 Robustness of Theorem 4

Detail-Freeness. Can the sender persuade the voters even when he does not know the exact details of the environment? We argue that Proposition 1 from the monopolistic sender setting holds in an even more general form when the voters hold exogenous private signals: here, to be able to persuade the voters, it is sufficient that the sender knows that Φ satisfies the monotonicity condition (1.21) and the richness assumption (1.17).

Specifically, the sender can release information to the voters such that, uniformly, for any prior $\Pr(\alpha) \in (0,1)$, any exogenous information π^c of the voters satisfying (1.20), and any aggregate preference function Φ satisfying (1.17) and (1.21), his target policy is implemented. This is possible simply by choosing the parameters of the general signal $\pi_n^{x,r,y}$ with x and y in $\{0,1\}$ and any arbitrary $r \in (0,1)$. Any such information structure implements a targeted policy uniformly as outlined before.

In a sense, the conditions for uniform implementability are weaker here than in Proposition 1 where we also required a condition on the prior. This is maybe

surprising if one thinks of the voters' exogenous information as a constraint on the sender. The reason it holds is that, with exogenous private information, the relevant "induced prior" after signal z, i.e., $\Pr(\alpha|\text{piv}, z)$, adjusts endogenously to ensure the equal-margin condition.

Basin of Attraction. The results from Section 1.4.6.2 regarding the basin of attraction of the manipulated equilibria for the case of a monopolistic sender do not extend when voters have exogenous private information.[18]

Other Equilibria. We conjecture that there always also exists a sequence of equilibria yielding the full-information outcome, as in the case of the monopolistic sender (see Theorem 3). However, we have not been able to prove this result so far for the case with private signals.

Belief Implementation. As in the monopolistic sender scenario, we provide a result more general than Theorem 4: In the spirit of the literature on Bayesian persuasion, we show that, for large electorates, there is a set of "implementable" posterior belief distributions, including arbitrarily extreme beliefs. This result is stated in Lemma 3 in the appendix, which is corresponds to Lemma 1 for the monopolistic sender. However, when voters have exogenous information, not all beliefs are implementable; this is consequential when the sender is only partially informed, discussed below in Section 1.6.1.

1.6 Remarks and Extensions

1.6.1 Partially Informed Sender

In the working paper version, Heese and Lauermann (2019), we consider a sender who does not know the state $\omega \in \{\alpha, \beta\}$.[19] Instead, the sender receives a private signal m. Conditional on the private signal m, the sender can release signals to the voters that are coarsenings of m.

Suppose that the sender's signal is binary, $m \in \{\ell, h\}$. Then, we show the following: if the sender is the monopolistic information provider (voters receive no private information), then the sender can implement any policy as a function of the own signal, i.e., for any $(x(\ell), x(h)) \in \{A, B\}^2$, the sender can ensure that a majority votes for $x(\ell)$ given the information released to voters after the own signal ℓ and for $x(h)$

18. Instead, one can show the following: let the sender release the information $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ to the voters as in Lemma 3. When the electorate is large enough, for almost any initial strategy, under the iterated best response, the voter behavior after signal z jumps back and forth indefinitely from voting approximately according to $\sigma^{\boldsymbol{p}}$, with $\boldsymbol{p} = \Pr(\alpha|s)_{s\in\{a,z,b\}}$, to voting approximately as if one of the states is known to be the true state. We omit the proof.

19. Available here https://ideas.repec.org/p/bon/boncrc/crctr224_2019_128.html.

after the own signal h. This is, in fact, implied by the analysis of the current paper. To see this, note that the sender's own signal m simply assumes the role of the state of nature ω in the current setting, and we can "integrate out" the state to rewrite the voters' preferences in terms of $\{\ell, h\}$.

However, when the voters have private information as well, the analysis is more subtle. Suppose voters observe an exogenous private signal π^c as in the CJT setting and the sender can release additional information in the form of a coarsening of the own noisy signal. For this case we show that, whenever the sender's own information is sufficiently precise relative to π^c, then again the sender can implement any policy as a function of the own signal, $(x(\ell), x(h)) \in \{A, B\}^2$; see Heese and Lauermann (2019, Theorem 7). For example, if the voters' signals $\{u, d\}$ are symmetric across states, then it is sufficient that the sender's own information is at least as informative as the joint signal of two voters (in the Blackwell sense).[20]

1.6.2 Known Preferences: Targeted Persuasion

When the types of the voters are known to a potential sender, voters can be "targeted" with recommendations; formally, a revelation principle applies saying that any equilibrium is equivalent to a recommendation policy that will be followed by the voters. Below, we show that when the preference types are known, there is a simple way in which the sender can persuade the voters to elect a constant policy via private recommendations.[21] We also show that, with known preferences, the possibility of persuasion is unaffected by the presence of a private signal of the voters.

Suppose that the voters' preference types $t^i = (t^i_\alpha, t^i_\beta)$ are commonly known for any $i \in \{1, \dots, 2n+1\}$. The voters receive exogenous private signals as in the setting of the CJT (Section 1.3.2) (the following result extends when these exogenous signals are uninformative). Suppose that the voters $1, \dots, m$ prefer A in α and B in β—that is $t^i_\alpha > 0$ and $t^i_\beta < 0$—and without loss let $m > n$. The remaining voters $m+1, \dots, 2n+1$ prefer B in α and A in β, that is $t^i_\alpha < 0$ and $t^i_\beta > 0$.[22]

The following recommendation policy implements the outcome A with probability of at least $1 - \epsilon$ in an equilibrium, for arbitrarily small $\epsilon > 0$: in both states, with probability $1 - \epsilon$, all voters receive the recommendation "vote A" (signal a). In state α, with the remaining probability ϵ, a random subset of size $n+1$ of the voters $1, \dots, m$ receives the recommendation "vote A" and the remaining n voters receive the recommendation "vote B" (signal b). In state β, with the remaining probabil-

20. This is shown in Heese and Lauermann (2019, Remark 2). The key step in the proof is the observation that, when the sender's signal is sufficiently precise, then the sender can induce beliefs that are "implementable" in the sense of Lemma 3 from the current appendix.

21. This has been observed by Chan et al. (2019) and in Bardhi and Guo (2018) in similar settings. Therefore, the main parts of these papers consider settings with voting costs ("expressive voting") and unanimity, respectively.

22. The example can be extended to include "partisans".

ity $\epsilon > 0$, a random subset of size $n+1$ of the voters $1, \ldots, m$ receives b and the remaining n voters receive a.

Voting A after an a-signal and B after a b-signal constitutes an equilibrium: given this strategy, denoted by σ, voters $i \in \{1, \ldots, m\}$ with an a-signal are only pivotal in α, and voters $i \in \{1, \ldots, m\}$ with a b-signal are only pivotal in β—that is $\Pr(\alpha|\text{piv}, a, i \leq m; \sigma) = 1$ and $\Pr(\alpha|\text{piv}, b, i \leq m; \sigma) = 0$. Hence, voting A after a and B after b is a strict best response for any voter $i \in \{1, \ldots, m\}$. Voters $i \in \{m+1, \ldots, 2n+1\}$ are never pivotal if the other voters follow the recommendations. Hence, following the recommendation is a best response also for them, and, therefore, σ is an equilibrium. Since with probability $1-\epsilon$ all citizens vote A, given σ, the recommendation policy implements the outcome A with a probability of at least $1-\epsilon$.

Note how the signal structure above is finely tuned to the details of the setting. By way of contrast, we show that persuasion is effective even if information can not be tailored to a specific preferences profile. In fact, we show that information does not even need to be tailored to the distribution of preferences. The mechanism driving persuasion is fundamentally different from the one described here. This difference may be most salient with exogenous private information where the equilibrium behavior of the voters adjusts endogenously to maintain the critical "equal-margin condition" across environments.

1.6.3 Bayes Correlated Equilibria

The Bayes correlated equilibria given some exogenous information structure π^c are the Bayes-Nash equilibria that arise from expansions π of π^c (see Bergemann and Morris (2016) for the definition of an expansion and the characterization of Bayes correlated equilibria). In terms of Bayes correlated equilibria, Theorem 4 means that for any state-dependent outcome function $(x(\alpha), x(\beta)) \in \{A, B\}^2$, there exists a sequence of Bayes correlated equilibria given π^c that leads to this outcome as $n \to \infty$.

1.7 Related Literature

Voter Persuasion Literature. The paper is related to work on information design in general (see Bergemann and Morris (2019) for a survey), especially with multiple receivers (e.g., Mathevet, Perego, and Taneva (2017)).

Previous work on persuasion in an election context has studied persuasion in settings in which the preferences of the voters are commonly known and voters have no access to exogenous private signals. The previous work has considered public signals by the sender (Alonso and Câmara, 2016), persuasion with conditionally independent private signals by the sender (Wang, 2013), and targeted persuasion with private signals by the sender (Bardhi and Guo, 2018; Chan, Gupta, Li, and

Wang, 2019). We discussed persuasion when the preferences of the voters are known in detail in Section 1.6.2, and we showed how the persuasion mechanism and its logic are quite different.

In contrast to the existing literature, we revisit the general voting setting of Feddersen and Pesendorfer (1997) with private preferences: in this setup, as a consequence of the Condorcet Jury Theorem, there is no scope for persuasion with public signals and also no scope for persuasion with conditionally independent private signals; see Theorem .

More generally, most of the Bayesian persuasion literature assumes that the sender has extensive knowledge of the environment; in particular, perfect knowledge about the state and the receiver's types is typically assumed.[23] In this paper, the informational requirements for persuasion are significantly weaker. We allow for private preferences and exogenous private signals of the receivers; we also consider the case when the sender has incomplete information regarding the prior probabilities of the state, the distribution of the private preference types of the voters, or the distribution of the private signals of the voters (see Section 1.4.6.1 and Section 1.5.4). In the working paper version, Heese and Lauermann (2019), we consider the case when the sender's information regarding the state is incomplete (see Section 1.6.1).

Several other papers study how groups can be influenced through strategic information transmission, but are less closely related: Kerman, Herings, and Karos (2019) study targeted persuasion via private signals when the sender is restricted to use signals that induce the voters to behave sincerely; compare to the discussion of targeted persuasion in Section 1.6.2. Levy, Moreno de Barreda, and Razin (2018) study persuasion of voters with correlation neglect. Schipper and Woo (2019) study the persuasion of voters who are unaware of certain features. Schnakenberg (2015) studies a cheap talk setting in which an expert tries to manipulate a voting body. Salcedo (2019) studies persuasion of subgroups of receivers via private messages in a setting where each receiver's payoff only depends on his own action and the state.

More distantly related is work on the design of an elicitation mechanism to elicit information from multiple experts for an adversary to use (Gerardi, McLean, and Postlewaite, 2009; Feng and Wu, 2019).

Information Aggregation Literature. Voting theory has identified several circumstances in which information may fail to aggregate. We discuss the studies that are most closely related: Feddersen and Pesendorfer (1997) (Section 6) show that an invertibility problem causes a failure when there is aggregate uncertainty with respect to the preference distribution conditional on the state. We have already mentioned

23. Exceptions are Guo and Shmaya (2019) and Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), who study persuasion of a single, privately informed receiver.

that Bhattacharya (2013) shows that information may fail to aggregate when preference monotonicity is violated.

In a pure common-values setting, Mandler (2012) shows that a failure can occur when there is aggregate signal uncertainty conditional on the state. There is a sense in which such aggregate uncertainty is necessary for a failure of information aggregation, in the sense that if there is a single substate, the CJT applies (Theorem 1 and the subsequent discussion). Here, as in his model, the voters' updating about the signal distribution of others conditional on a close election is important.[24] Note that the pure common-values assumption implies that this setting is a special case of a setting where the individual voters' preference type is known, discussed in Section 1.6.2. In contrast to Mandler (2012), we consider a setting in which voters do not have common values and study the effect of an additional signal in the canonical setting by Feddersen and Pesendorfer (1997) rather than perturbing that original signal.

Further related models of elections that perform poorly in aggregating information are, among others, Razin (2003), Acharya (2016), Ekmekci and Lauermann (2019), Ali, Mihm, and Siga (2018) and Bhattacharya (2018).

1.8 Conclusion

In the canonical voting setting by Feddersen and Pesendorfer (1997), information aggregation may be upset by an interested sender who provides additional information to the voters. We have shown how an interested sender can exploit strategic voters by manipulating their inference from the election being close. The sender does not need precise knowledge of the environment ("detail-freeness"), and the same information structure is effective uniformly across model specifications. In fact, the same information structure that implements a given policy in the monopolistic sender setting also implements the policy when voters have private information. Even a manipulator with very limited knowledge about the state itself can persuade a large electorate, as we show in the working paper version, Heese and Lauermann (2019). When the sender is the monopolistic information provider, we demonstrated the effectiveness of persuasion in a small election with just 15 voters. We also showed that the resulting equilibrium is simple and selected by an iterated best response dynamic.

The pivotal voting model considers the extreme case where voters react perfectly to the closeness of the election when interpreting their information and illustrates the effectiveness of persuasion in this case. One may conjecture that, in a setting in

24. Uncertainty regarding the signal distribution and updating about it is also central in Acharya and Meirowitz (2017), where aggregate uncertainty supports sincere voting. Other recent contributions on the conditions for information aggregation are Kosterina (2019) and Barelli, Bhattacharya, and Siga (2019).

which voters react less sensitively, persuasion is still effective but, presumably, less so—a conjecture that may be worthwhile exploring.

Conceptually, our results also mean that equilibrium outcomes in the setting by Feddersen and Pesendorfer (1997) can be hard to predict for an outside observer without precise knowledge of the voters' information. The outside observer must be able to exclude the possibility that voters have access to additional information of the form discussed here.

Finally, information aggregation has also been studied in (double-) auctions, a setting that shares some features with elections. An interesting question may be whether, in auctions, information aggregation is an "informationally robust" prediction or whether bidders having additional information can also upset it. Information design in auction settings has been studied, among others, by Bergemann, Brooks, and Morris (2016), Du (2018), and Yamashita et al. (2016) but mostly with a focus on revenue and efficiency.

1.A Proof of the Condorcet Jury Theorem

Step 1. For all n and every equilibrium σ_n^*, the vote share of A is larger in α than in β,

$$0 < q(\beta; \sigma_n^*, n) < q(\alpha; \sigma_n^*, n) < 1. \tag{1.65}$$

This ordering of the vote shares follows from the likelihood ratio ordering of the signals. In particular, recall the expression (1.8) for the posterior likelihood ratio of two states conditional on a given voter's signal s and the event that the voter is pivotal,

$$\frac{\Pr(\alpha|s, \text{piv}; \sigma_n^*, n)}{1 - \Pr(\alpha|s, \text{piv}; \sigma_n^*, n)} = \frac{\Pr(\alpha)}{\Pr(\beta)} \frac{\Pr(\text{piv}|\alpha; \sigma_n^*, n)}{\Pr(\text{piv}|\beta; \sigma_n^*, n)} \frac{\Pr(s|\alpha; \pi^c)}{\Pr(s|\beta; \pi^c)}, \tag{1.66}$$

where $\Pr(\text{piv}|\beta; \sigma_n^*, n) > 0$ because σ_n^* is nondegenerate by (1.2). Therefore, $\frac{\Pr(u|\alpha;\pi^c)}{\Pr(u|\beta;\pi^c)} > \frac{\Pr(d|\alpha;\pi^c)}{\Pr(d|\beta;\pi^c)}$ implies that $\Pr(\alpha|u, \text{piv}; \sigma_n^*, n) > \Pr(\alpha|d, \text{piv}; \sigma_n^*, n)$. Now, (1.65) follows from (1.15) and the monotonicity of Φ. Intuitively, the expected posterior in state α is higher and this translates into a larger set of types preferring A given the monotonicity of Φ.

Step 2. Voters cannot become certain of the state conditional on being pivotal, that is, the inference from the pivotal event must remain bounded,

$$\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha; \sigma_n^*, n)}{\Pr(\text{piv}|\beta; \sigma_n^*, n)} \in (0, \infty), \tag{1.67}$$

for every convergent subsequence in the extended reals.

Suppose not and suppose instead, for example, that conditional on being pivotal, voters become convinced that the state is β, i.e., $\eta = \lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha;\sigma_n^*,n)}{\Pr(\text{piv}|\beta;\sigma_n^*,n)} = 0$. This would imply $\lim_{n\to\infty} \Pr(\alpha|s, \text{piv}; \sigma_n^*, n) = 0$ for $s \in \{u, d\}$. Then, given $\Phi(0) < \frac{1}{2}$, a strict majority would support B in both states. However, the election is then closer to being tied in state α and voters would update toward state α conditional on being pivotal, in contradiction to $\eta = 0$.

Formally, if $\eta = 0$ for some converging subsequence, then $\lim_{n\to\infty} q(\omega; \sigma_n^*) = \Phi(0) < \frac{1}{2}$ for $\omega \in \{\alpha, \beta\}$. Therefore, for large enough n, (1.65) implies that $q(\beta; \sigma_n^*) < q(\alpha; \sigma_n^*) < 1/2$. Now, Claim 1 implies that voters update toward state α, that is, $\frac{\Pr(\text{piv}|\alpha;\sigma_n^*,n)}{\Pr(\text{piv}|\beta;\sigma_n^*,n)} \geq 1$, in contradiction to $\eta = 0$.

Step 3. In every equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$, the limit of the vote share of A is larger in α than in β,

$$\lim_{n\to\infty} q(\alpha; \sigma_n^*) > \lim_{n\to\infty} q(\beta; \sigma_n^*). \tag{1.68}$$

From (1.67) and (1.66), we have that the limits of the posteriors conditional on being pivotal and $s \in \{u, d\}$ are interior and hence ordered,

$$0 < \lim_{n\to\infty} \Pr(\alpha|d, \text{piv}; \sigma_n^*, n) < \lim_{n\to\infty} \Pr(\alpha|u, \text{piv}; \sigma_n^*, n) < 1.$$

Now, (1.68) follows from (1.15) since Φ is strictly increasing.

Step 4. The election is equally close to being tied in expectation, that is, (1.22) holds:

$$\lim_{n\to\infty} q(\alpha;\sigma_n^*) - \frac{1}{2} = \lim_{n\to\infty} \frac{1}{2} - q(\beta;\sigma_n^*).$$

Since voters must not become certain conditional on being pivotal by (1.67), Claim 2 requires that

$$\lim_{n\to\infty} \left| q(\alpha;\sigma_n^*) - \frac{1}{2} \right| = \lim_{n\to\infty} \left| q(\beta;\sigma_n^*) - \frac{1}{2} \right|. \tag{1.69}$$

Given the ordering of the limits of the vote shares from (1.68), the equation (1.69) implies (1.22).

It follows from Step 4 and (1.68) that

$$\lim_{n\to\infty} q(\alpha;\sigma_n^*) > \frac{1}{2} > \lim_{n\to\infty} q(\beta;\sigma_n^*).$$

Therefore, by the weak law of large numbers, A wins in state α with probability converging to 1 as $n \to \infty$ and B wins in state β with probability converging to 1 as $n \to \infty$. This proves Theorem 1.

Sketch of the proof of Theorem '. To see why the theorem is true, note that, given the binary state, the signals can be taken to be ordered by the monotone likelihood ratio, without loss of generality. For any fixed information structure π and any equilibrium σ_n^*, it then follows from (1.66) that the distribution of posteriors $\Pr(\alpha|\text{piv}, s; \sigma_n^*, \pi, n)$ in the state α (as implied by the distribution over s) first order stochastically dominates the distribution of posteriors $\Pr(\alpha|\text{piv}, s; \sigma_n^*, \pi, n)$ in the state β. Then, given that Φ is monotone, it follows from (1.15) that the vote shares satisfy the ordering (1.65). From (1.65) onward none of the arguments use that the signals are binary.

By the same line of argument, Theorem 1 holds even when we allow the information structure π_n with a single substate to vary with n (keeping the signal set S fixed), as long as the limit information structure is not completely uninformative.

1.B Monopolistic Persuasion

1.B.1 Proof of Claim 3

Without loss of generality, suppose σ_n is such that $q(\alpha_2;\sigma_n)(1-q(\alpha_2;\sigma_n)) < q(\beta_2;\sigma_n)(1-q(\beta_2;\sigma_n))$ for all n. It follows directly from (1.7) that

$$\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha_2;\sigma_n,\pi_n)}{\Pr(\text{piv}|\beta_2;\sigma_n,\pi_n)} \leq 1. \tag{1.70}$$

We now show that the reverse inequality also holds and thereby finish the proof of the lemma. For this, we show the following. There exists some $L > 0$ and $M > 0$ such that, for all n and all σ_n satisfying the ordering above,

$$\frac{\Pr(\text{piv}|\alpha_2;\sigma_n,\pi_n)}{\Pr(\text{piv}|\beta_2;\sigma_n,\pi_n)} \geq \left(1-\frac{L}{Mn^2}\right)^n. \tag{1.71}$$

First, it follows from (1.15) that the expected vote share for A in α_2 differs from the expected vote share for A in β_2 maximally by the probability that b is observed in α_2, that is, by $\varepsilon^2 = \frac{1}{n^2}$; so,

$$|q(\alpha_2;\sigma_n) - q(\beta_2;\sigma_n)| \leq \epsilon^2, \tag{1.72}$$

for all n. Second, recall that $\Phi(0) < q(\omega_j;\sigma) < \Phi(1)$ for any strategy and any substate ω_j, and note that the derivative of $h(q) = q(1-q)$ is bounded by some $L > 0$ on the compact interval $[\Phi(0), \Phi(1)]$. These observations taken together imply that

$$h(q(\beta_2;\sigma_n))\left|\frac{h(q(\alpha_2;\sigma_n))}{h(q(\beta_2;\sigma_n)} - 1\right| = |h(q(\alpha_2;\sigma_n)) - h(q(\beta_2;\sigma_n))| \leq L\epsilon^2. \tag{1.73}$$

for all n. Since $0 < \Phi(0) < q(\alpha_2;\sigma_n) < \Phi(1)$ and h is inverse U-shaped with maximum at $\frac{1}{2}$, this bound implies

$$\frac{h(q(\alpha_2;\sigma_n))}{h(q(\beta_2;\sigma_n)} \geq 1 - \frac{L}{h(q(\beta_2;\sigma_n))n^2} \geq 1 - \frac{L}{Mn^2} \tag{1.74}$$

for $M = \min(h(\Phi(0)), h(\Phi(1)))$ and all n. Now, (1.71) follows from (1.7).

Finally, since $\lim_{n\to\infty}(1-\frac{L}{Mn^2})^n = 1$, (1.71) implies that

$$\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha_2;\sigma_n,\pi_n)}{\Pr(\text{piv}|\beta_2;\sigma_n,\pi_n)} \geq 1. \tag{1.75}$$

To see why $\lim_{n\to\infty}(1-\frac{L}{Mn^2})^n = 1$, note that $\lim_{n\to\infty}(1-\frac{L}{Mn^2})^{2n} = \lim_{n\to\infty}(1-\frac{\sqrt{L}}{\sqrt{M}n})^{2n}(1+\frac{\sqrt{L}}{\sqrt{M}n})^{2n} = e^{2\sqrt{\frac{L}{M}}}e^{-2\sqrt{\frac{L}{M}}} = e^0 = 1$ where we used $\lim_{n\to\infty}(1+\frac{x}{n})^n = e^x$. This finishes the proof of Claim 3.

1.B.2 Computational Example

Note that one example of a distribution G on $[0,1]\times[-1,0]$ that induces a uniform distribution of 'thresholds of doubt', i.e. Φ with $\Phi(p) = p$ for all $p \in [0,1]$ is given by the density[25]

$$g(t_\alpha, t_\beta) = \begin{cases} \sqrt{1+(\frac{t_\beta}{t_\alpha})^2}^{-1} \cdot (2\cdot\int_{|t_\alpha|>|t_\beta|}\sqrt{1+(\frac{t_\beta}{t_\alpha})^2}^{-1}dt)^{-1} & \text{if} \quad \frac{-t_\beta}{t_\alpha - t_\beta} \leq \frac{1}{2}, \\ \sqrt{1+(\frac{t_\alpha}{t_\beta})^2}^{-1} \cdot (2\cdot\int_{|t_\alpha|>|t_\beta|}\sqrt{1+(\frac{t_\beta}{t_\alpha})^2}^{-1}dt)^{-1} & \text{if} \quad \frac{-t_\beta}{t_\alpha - t_\beta} \geq \frac{1}{2}. \end{cases}$$

We utilize the following auxiliary result.

25. To see why, note that for each $t >> 0$, $d(t) = \left[\sqrt{1+(\frac{t_\alpha}{t_\beta})^2}\right]$ is the length of the indifference plane of t. By setting the density of types proportional to $\frac{1}{d(t)}$, integrating over each indifference plane gives the same number such that types are uniformly distributed across indifference planes.

Lemma 2. Consider any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ and any sequence of information structures $(\pi_n)_{n\in\mathbb{N}}$ with a common set of substates across n. Then, for any substates $\omega_i, \omega_j' \in \{\alpha_1, \dots, \alpha_{N_\alpha}\} \cup \{\beta_1, \dots, \beta_{N_\beta}\}$ and any $n \in \mathbb{N}$,

$$\frac{\Pr(\text{piv}|\omega_i; \sigma_n, \pi_n)}{\Pr(\text{piv}|\omega_j'; \sigma_n, \pi_n)} = \Big[1 + \frac{(q(\omega_j'; \sigma^{\boldsymbol{p}}) - \frac{1}{2})^2 - (q(\omega_i; \sigma^{\boldsymbol{p}}) - \frac{1}{2})^2}{\frac{1}{4} - (q(\omega_j'; \sigma^{\boldsymbol{p}}) - \frac{1}{2})^2}\Big]^n \tag{1.76}$$

Proof. Let $x_n = q(\omega_i; \sigma_n) - \frac{1}{2}$ and $y_n = q(\omega_j'; \sigma_n) - \frac{1}{2}$. Then,

$$\begin{aligned}
\frac{q(\omega_i; \sigma_n)(1 - q(\omega_i; \sigma_n))}{q(\omega_j'; \sigma_n)(1 - q(\omega_j'; \sigma_n))} &= \frac{(\frac{1}{2} + x_n)(\frac{1}{2} - x_n)}{(\frac{1}{2} + y_n)(\frac{1}{2} - y_n)} \\
&= \frac{\frac{1}{4} - y_n^2 + y_n^2 - x_n^2}{\frac{1}{4} - y_n^2} \\
&= 1 + \frac{y_n^2 - x_n^2}{\frac{1}{4} - y_n^2}
\end{aligned}$$

The claim follows from (1.8). □

Fixed Point Argument.
Consider a belief $\boldsymbol{p} = (p_a, p_z, p_b)$ with

$$p_a \geq 0.95, \tag{1.77}$$
$$p_b \geq 0.95, \tag{1.78}$$
$$p_z \in [0.32, 0.68]. \tag{1.79}$$

Given $(\pi_n)_{n\in\mathbb{N}} = (\pi_n^r)_{n\in\mathbb{N}}$ with $r = \frac{1}{2}$, we have the following bounds for $n \geq 8$:

$$q(\omega_1; \sigma^{\boldsymbol{p}}, n) \geq 0.95 \quad \text{for} \quad \omega_1 \in \{\alpha_1, \beta_1\}, \tag{1.80}$$
$$q(\alpha_2; \sigma^{\boldsymbol{p}}, n) > 0.3 \tag{1.81}$$
$$q(\beta_2; \sigma^{\boldsymbol{p}}, n) \leq 0.7. \tag{1.82}$$

In the following, we omit the dependence on $\sigma^{\boldsymbol{p}}$ and on π_n most of the time.

Step 1. For any $n \in \mathbb{N}$ and any $\omega_1 \in \{\alpha_1, \beta_1\}, \omega_2' \in \{\alpha_2, \beta_2\}$,

$$\frac{\Pr(\text{piv}|\omega_2')}{\Pr(\text{piv}|\omega_1)} \geq (4.4)^n \tag{1.83}$$

Indeed,

$$\begin{aligned}
&\frac{\Pr(\text{piv}|\omega_2')}{\Pr(\text{piv}|\omega_1)} \\
\geq\ & [1+\min_{\omega_1,\omega_2'}\frac{(q(\omega_1;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2-(q(\omega_2';\sigma^{\boldsymbol{p}})-\frac{1}{2})^2}{\frac{1}{4}-(q(\omega_1;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2}]^n \\
\geq\ & (1+(\frac{(\frac{9}{20})^2-(\frac{4}{20})^2}{\frac{1}{4}-(\frac{9}{20})^2})) \\
\geq\ & (1+\frac{65}{19})^n \\
\geq\ & (4.4)^n. \qquad (1.84)
\end{aligned}$$

where we used Lemma 2 for the inequality on the second line.

Step 2. For $n \geq 7$: $\rho_a(\sigma^{\boldsymbol{p}}) \geq 0.95$, $\rho_b(\sigma^{\boldsymbol{p}}) \geq 0.95$ and $\rho_z(\sigma^{\boldsymbol{p}}) \in [0.32, 0.68]$.

First,

$$\rho_a(\sigma^{\boldsymbol{p}}) = 1 \qquad (1.85)$$

since a is only sent in α. Second,

$$\begin{aligned}
\frac{\rho_b(\sigma^{\boldsymbol{p}})}{1-\rho_b(\sigma^{\boldsymbol{p}})} &= \frac{p_0}{1-p_0}\frac{\Pr(\alpha_2|\alpha)\Pr(b|\alpha_2)\Pr(\text{piv}|\alpha_2)}{\Pr(\beta_1|\beta)\Pr(b|\beta_1)\Pr(\text{piv}|\beta_1)} \\
&\geq \frac{\frac{1}{n}\frac{1}{n^2}}{(1-\frac{1}{n})}(4.4)^n \\
&\geq 100 \quad \text{for} \quad n \geq 7.
\end{aligned}$$

where we used (1.84) for the inequality on the second line. Hence, for $n \geq 7$,

$$\rho(\sigma^{\boldsymbol{p}})_b \geq \frac{100}{101} > 0.95. \qquad (1.86)$$

Third,

$$\begin{aligned}
\frac{\Pr(\text{piv}|\alpha_2)}{\Pr(\text{piv}|\beta_2)} &\leq [1+\frac{|(q(\beta_2;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2-(q(\alpha_2;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2|}{\frac{1}{4}-(q(\beta_2;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2}]^n \\
&\leq (1+\frac{\frac{1}{n^4}+\frac{1}{n^2}}{\frac{1}{4}-\frac{16}{400}})^n \\
&\leq 2. \quad \text{for} \quad n \geq 7.
\end{aligned}$$

where we used Lemma 2 for the inequality on the first line. For the inequality on the second line, we used that z is sent with probability $1-\frac{1}{n^2}$ in both α_2 and β_2 such that the difference in the squared margins of victory cannot exceed $(x+\frac{1}{n^2})^2 - x^2 \leq \frac{2x}{n^2}+\frac{1}{n^4}$ where x is the minimum margin of victory in the states α_2, β_2. Finally, the inequality follows since the margin of victory in both α_2 and β_2 is bounded by 0.2.

So,

$$\begin{aligned}\frac{\rho_z(\sigma^{\boldsymbol{p}})_z}{1-\rho_z(\sigma^{\boldsymbol{p}})} &= \frac{\Pr(\alpha)}{\Pr(\beta)}\frac{\Pr(\alpha_2|\alpha)}{\Pr(\beta_2|\beta)}\frac{\Pr(z|\alpha_2)}{\Pr(z|\beta_2)}\frac{\Pr(\text{piv}|\alpha_2)}{\Pr(\text{piv}|\beta_2)}\\ &= (1-\frac{1}{n^2})\frac{\Pr(\text{piv}|\alpha_2)}{\Pr(\text{piv}|\beta_2)}\\ &\leq 2 \quad \text{for} \quad n \geq 7.\end{aligned}$$

Consequently, for all $n \geq 7$,

$$\rho(\sigma^{\boldsymbol{p}})_z \leq \frac{2}{3}. \tag{1.87}$$

Fourth,

$$\begin{aligned}\frac{\Pr(\text{piv}|\alpha_2)}{\Pr(\text{piv}|\beta_2)} &\geq (1-\frac{|(q(\beta_2;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2-(q(\alpha_2;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2|}{\frac{1}{4}-(q(\beta_2;\sigma^{\boldsymbol{p}})-\frac{1}{2})^2}\\ &\geq (1-\frac{\frac{1}{n^4}+\frac{1}{n^2}}{\frac{1}{4}-\frac{16}{400}})^n\\ &\geq 0.48 \quad \text{for} \quad n \geq 7. \end{aligned}\tag{1.88}$$

So, for all $n \geq 7$,

$$\begin{aligned}\frac{\rho(\sigma^{\boldsymbol{p}})_z}{1-\rho(\sigma^{\boldsymbol{p}})_z} &= (1-\frac{1}{n^2})\frac{\Pr(\text{piv}|\alpha_2;\sigma^{\boldsymbol{p}})}{\Pr(\text{piv}|\beta_2;\sigma^{\boldsymbol{p}})}\\ &\geq 0.471.\end{aligned}$$

This gives for all $n \geq 7$,

$$\rho(\sigma^{\boldsymbol{p}})_z \geq \frac{0.471}{1+0.471} \geq 0.32. \tag{1.89}$$

The claim follows from (1.85) - (1.89).

Step 3. For $n \geq 7$, there is an equilibrium σ_n^* which satisfies (1.80) - (1.82).

It follows from Step 2 that, for any $n \geq 7$, the continuous map that sends $\boldsymbol{p}$ to $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ is a self-map on the set of beliefs that satisfy (1.77) - (1.79). It follows from the Kakutani fixed point theorem that there exists fixed points $\boldsymbol{p}_n^*$ that satisfy (1.77) - (1.79). The corresponding strategies $\sigma^{\boldsymbol{p}_n^*}$ are equilibria (compare to (1.13)) and they satisfy (1.80) - (1.82).

Step 4. Given the equilibrium σ_n^* for $n \geq 7$, the probability that A is elected is larger than $99.9\% \cdot (1-\frac{1}{n})$.

Evaluation of the binomial distribution shows that $\Pr(\mathscr{B}(2n+1,x)) > n) \geq 0.999999$ if $n \geq 7$ and $x \geq 0.95$. Hence, given σ_n^*, A is elected with probability larger than 99.9% in the states α_1 and β_1. Finally, the claim follows since these states occur with probability larger than $(1-\frac{1}{n})$. The fourth step finishes the calculations for the example.

1.B.3 Proof of Lemma 1

1.B.3.1 Preliminaries: Voter Inference

The basic arguments of the previous discussion of the voters' inference from Section 1.4.2.2 extend to the general case.

Consider the signal z and the inference about the relative likelihood of α_2 and β_2. As in Claim 3, for *any* strategy used by the other voters, the pivotal event contains no information about the relative probability of α_2 and β_2 as the electorate grows large.

Claim 7. Given any parameters $(x,r,y) \in [0,1]^3$ and any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$,

$$\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha_2;\sigma_n,\pi_n^{x,r,y})}{\Pr(\text{piv}|\beta_2;\sigma_n,\pi_n^{x,r,y})} = 1. \tag{1.90}$$

The arguments from the proof of the analogous Claim 3 hold verbatim with the required changes in notation; therefore, the proof is omitted. Claim 7 and (1.42) imply, in particular, that

$$\lim_{n\to\infty} \frac{\Pr(\alpha|z,\text{piv};\sigma_n,\pi_n^{x,r,y})}{\Pr(\beta|z,\text{piv};\sigma_n,\pi_n^{x,r,y})} = \frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r}. \tag{1.91}$$

Next, we consider a signal $s \in \{a,b\}$ and the voters' inference about the relative likelihood of α and β. We show that, analogous to Claim 4, for this signal, the inference from the signal is dominated by the inference from being pivotal if the election is closer to being tied in states α_2 and β_2 than in the states α_1 and β_1.

Claim 8. Take any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ such that

$$\begin{aligned} &\lim_{n\to\infty} \max_{\omega_2\in\{\alpha_2,\beta_2\}} |q(\sigma_n;\omega_2,\pi_n^{x,r,y}) - \frac{1}{2}| \\ <\ &\lim_{n\to\infty} \min_{\omega_1\in\{\alpha_1,\beta_1\}} |q(\sigma_n;\omega_1,\pi_n^{x,r,y}) - \frac{1}{2}|; \end{aligned} \tag{1.92}$$

then, for $s \in \{a,b\}$,

$$\lim_{n\to\infty} \frac{\Pr(\{\alpha_2,\beta_2\}|s,\text{piv};\sigma_n,\pi_n^{x,r,y})}{\Pr(\{\alpha_1,\beta_1\}|s,\text{piv};\sigma_n,\pi_n^{x,r,y})} = \infty. \tag{1.93}$$

The claim follows from the same arguments as Claim 4, and we omit this proof as well.

For any sequence of strategies that satisfies (1.92), Claims 7 and 8 imply that, for signal a,[26]

26. Recall the convention $\frac{1}{0} = \infty$, such that, for $x = 1$, the following equalities hold in the extended reals.

$$\begin{aligned}\lim_{n\to\infty}\frac{\Pr(\alpha|a,\text{piv};\sigma_n,\pi_n^{x,r,y})}{\Pr(\beta|a,\text{piv};\sigma_n,\pi_n^{x,r,y})} &= \frac{\Pr(\alpha_2|\{\alpha_2,\beta_2\},a;\sigma_n,\pi_n^{x,r,y})}{\Pr(\beta_2|\{\alpha_2,\beta_2\},a;\sigma_n,\pi_n^{x,r,y})}\\ &= \frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r}\frac{x}{1-x} \end{aligned} \tag{1.94}$$

and that for signal b,

$$\begin{aligned}\lim_{n\to\infty}\frac{\Pr(\alpha|b,\text{piv};\sigma_n,\pi_n^{x,r,y})}{\Pr(\beta|b,\text{piv};\sigma_n,\pi_n^{x,r,y})} &= \frac{\Pr(\alpha_2|\{\alpha_2,\beta_2\},b;\sigma_n,\pi_n^{x,r,y})}{\Pr(\beta_2|\{\alpha_2,\beta_2\},b;\sigma_n,\pi_n^{x,r,y})}\\ &= \frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r}\frac{y}{1-y}. \end{aligned} \tag{1.95}$$

1.B.3.2 Implementable Beliefs

We use that an equilibrium is equivalently characterized by a vector of beliefs, $\boldsymbol{p}^* = (p_a^*, p_z^*, p_b^*)$ such that $\boldsymbol{p}^* = \boldsymbol{\rho}\left(\sigma^{\boldsymbol{p}^*}\right)$; see (1.13). Take any $\delta > 0$ and let

$$\mathrm{B}_\delta = \left\{\boldsymbol{p} \in [0,1]^3 \mid |\boldsymbol{p} - (\mu_\alpha, r', \mu_\beta)| \leq \delta\right\}, \tag{1.96}$$

so that B_δ is the set of beliefs at most δ away from $(\mu_\alpha, r', \mu_\beta)$.

We show that Claim 7 and 8 imply that there is a large set of belief triples $(\mu_\alpha, r', \mu_\beta)$ such that, given $\sigma^{\boldsymbol{p}}$, the posterior conditional on being pivotal is again in B_δ, for any $\boldsymbol{p} \in \mathrm{B}_\delta$, any sufficiently small δ and any sufficiently large n.[27]

Claim 9. Let $(\mu_\alpha, \mu_\beta) \in [0,1]^2$ and $r' \in (0,1)$ with

$$|\Phi(\mu_\alpha) - \frac{1}{2}| > |\Phi(r') - \frac{1}{2}| \text{ and } |\Phi(\mu_\beta) - \frac{1}{2}| > |\Phi(r') - \frac{1}{2}|. \tag{1.97}$$

For any $\delta > 0$ small enough, there exists $n(\delta)$ such that for all $n \geq n(\delta)$,

$$\forall \boldsymbol{p} \in B_\delta : \boldsymbol{\rho}\left(\sigma^{\boldsymbol{p}}; \pi_n^{x,r,y}, n\right) \in B_\delta \tag{1.98}$$

for (x,r,y) being the solutions to $\frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r}\frac{x}{1-x} = \frac{\mu_\alpha}{1-\mu_\alpha}$, $\frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r}\frac{y}{1-y} = \frac{\mu_\beta}{\mu_\beta}$, and $\frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r} = \frac{r'}{1-r'}$.

Proof. Let $\pi_n = \pi_n^{x,r,y}$. Take any $\boldsymbol{p} \in B_\delta$ and consider the corresponding strategy $\sigma^{\boldsymbol{p}}$. The condition (1.97) implies that for δ small enough, the election is closer to being tied in the states α_2 and β_2 than in the states α_1 and β_1 in expectation as $n \to \infty$:

$$\begin{aligned}\forall \boldsymbol{p} \in \mathrm{B}_\delta: \qquad & \lim_{n\to\infty} \max_{\omega_2 \in \{\alpha_2,\beta_2\}} |q(\omega_2; \sigma^{\boldsymbol{p}}, \pi_n) - \frac{1}{2}|\\ < \quad & \lim_{n\to\infty} \min_{\omega_1 \in \{\alpha_1,\beta_1\}} |q(\omega_1; \sigma^{\boldsymbol{p}}, \pi_n) - \frac{1}{2}|. \end{aligned} \tag{1.99}$$

27. In the following, we use the convention that dividing by zero yields a result of infinity such that formulas like $\frac{\Pr(\alpha)}{\Pr(\beta)}\frac{r}{1-r}\frac{x}{1-x} = \frac{\mu_\alpha}{1-\mu_\alpha}$ make sense for $\mu_\alpha \in \{0,1\}$.

To see why, note that for n large enough, $q(\alpha_2;\sigma^{\boldsymbol{p}},\pi_n)\approx\Phi\left(p_z\right)$ and $q(\beta_2;\sigma^{\boldsymbol{p}},\pi_n)\approx\Phi\left(p_z\right)$ since almost all voters receive z in α_2 and β_2. Also, $q(\alpha_1;\sigma^{\boldsymbol{p}},\pi_n)=\Phi\left(p_a\right)$ since all voters receive a in α_1 and $q(\beta_1;\sigma^{\boldsymbol{p}},\pi_n)=\Phi\left(p_b\right)$ since all voters receive b in β_1. In addition, by the continuity of Φ, for δ small enough, we have that $\Phi\left(p_z\right)\approx\Phi\left(r'\right)$, $\Phi\left(p_a\right)\approx\Phi\left(\mu_\alpha\right)$ and $\Phi\left(p_b\right)\approx\Phi\left(\mu_\beta\right)$. Finally, (1.99) follows then from $\Phi\left(\hat{r}\right)=\frac{1}{2}$ and $\Phi\left(\mu_\omega\right)\neq\frac{1}{2}$ for $\omega\in\{\alpha,\beta\}$. Now, it follows from (1.99), Claim 8, and its implications (1.94) and (1.95) that

$$\lim_{n\to\infty}\boldsymbol{\rho}_a\left(\sigma^{\boldsymbol{p}};\pi_n,n\right) = \mu_\alpha, \tag{1.100}$$

$$\lim_{n\to\infty}\boldsymbol{\rho}_b\left(\sigma^{\boldsymbol{p}};\pi_n,n\right) = \mu_\beta. \tag{1.101}$$

for any $\delta>0$ small enough. Thus, the claim follows from (1.91), (1.100) and (1.101). □

We finish the proof of Lemma 1. Let $r=\frac{\Pr(\alpha)\hat{r}}{\Pr(\alpha)\hat{r}+(1-\Pr(\alpha))\left(1-\hat{r}\right)}$ with $\Phi(\hat{r})=\frac{1}{2}$; see (1.40). Take any (μ_α,μ_β) with $\Phi(\mu_\alpha)\neq\frac{1}{2}$ and $\Phi(\mu_\beta)\neq\frac{1}{2}$. Then, given Claim 9, $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ is a self-map on B_δ for δ small enough and $n\geq n(\delta)$. Since $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ is continuous in $\boldsymbol{p}$, it follows from Kakutani's theorem that there exists a fixed point $\boldsymbol{p}_n^*\in B_\delta$ for all n large enough, i.e., $\boldsymbol{p}_n^*=\boldsymbol{\rho}(\sigma^{\boldsymbol{p}_n^*})$ and the corresponding behavior $\sigma^{\boldsymbol{p}_n^*}$ forms a sequence of equilibria. Lemma 1 follows from (1.100) and (1.101).

1.B.4 Proof of Proposition 1

We provide the proof for the constant target policy A in both states, i.e., $(x(\alpha),x(\beta))=(A,A)$. Let the sender use the information structures $\pi_n=\pi_n^{x,r,y}$ with $x=y=1$ and $r=\frac{1}{2}$. It follows from Claim 9 that, for any Φ for which (1.49) and (1.50) hold, there is a δ small enough such that $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ is a self-map on $B_\delta=\{\boldsymbol{p}\in[0,1]^3:|\boldsymbol{p}-(1,\Pr(\alpha),1)|\leq\delta\}$ for all n large enough.
Since $\boldsymbol{\rho}(\sigma^{\boldsymbol{p}})$ is continuous in $\boldsymbol{p}$, it follows from Kakutani's theorem that there exists a fixed point $\boldsymbol{p}_n^*\in B_\delta$ for all n large enough, i.e., $\boldsymbol{p}_n^*=\boldsymbol{\rho}(\sigma^{\boldsymbol{p}_n^*})$ and the corresponding behavior $\sigma^{\boldsymbol{p}_n^*}$ forms a sequence of equilibria that implements the beliefs $(\mu_\alpha,\mu_\beta)=(1,1)$. Given $(\sigma^{\boldsymbol{p}_n^*})_{n\in\mathbb{N}}$, the policy A wins in both states; this follows since voters with an a and b-signal are supporting A with a probability converging to $\Phi(1)>\frac{1}{2}$ and from the weak law of large numbers. The other cases are analogous. This finishes the proof of the lemma.

1.B.5 Proof of Proposition 2 (Basin of Attraction)

Recall that for any strategy σ, the distance between the margin of victory in α_2 and β_2 is smaller than $\frac{2}{n^2}$ in expectation since the probability that a voter receives the signal z is at least $1-\frac{2}{n^2}$ in both the substates. Now, consider any belief $\boldsymbol{p}\in[0,1]^3$ such that under the corresponding strategy $\sigma^{\boldsymbol{p}}$ the margins of victory differ by at least $\delta>0$ for any other pair of substates. The theorem follows from the following

claim: we show that for any such belief $\boldsymbol{p}$, the twice iterated response is δ-close to the manipulated equilibrium when n is large enough.

Claim 10. Take any beliefs $(\mu_\alpha, \mu_\beta) \in [0,1]^2$ with $\Phi(\mu_\alpha) \neq \frac{1}{2}$ and $\Phi(\mu_\beta) \neq \frac{1}{2}$ and the corresponding information structures $(\pi_n^{x,r,y})$ from Lemma 1.

For any $\delta > 0$, there exists $\bar{n} \in \mathbb{N}$ s.t., for any $\boldsymbol{p} \in [0,1]^3$ for which

$$\left| |q(\omega_i, \sigma^{\mathbf{p}}, \pi_n) - \frac{1}{2}| - |q(\omega_j', \sigma^{\boldsymbol{p}}, \pi_n) - \frac{1}{2}| \right| > \delta, \tag{1.102}$$

for all $\omega_i \in \{\alpha_1, \alpha_2, \beta_1, \beta_2\}$ and $\omega_j' \in \{\alpha_1, \beta_1\}$ with $\omega_i \neq \omega_j'$, it holds that, for $n \geq \bar{n}$,

$$|\boldsymbol{\rho}^2(\boldsymbol{p}) - (\mu_\alpha, \hat{r}, \mu_\beta)| < \delta. \tag{1.103}$$

The claim implies Proposition 2 because δ can be chosen arbitrarily small.

Proof. Take any $\boldsymbol{p} \in [0,1]^3$ such that (1.102) holds and consider the corresponding behavior $\sigma^{\boldsymbol{p}}$. Denote the best response to $\sigma^{\boldsymbol{p}}$ by $\tilde{\sigma} = \sigma^{\boldsymbol{\rho}(\sigma^{\boldsymbol{p}}; \pi_n, n)}$ and let $\pi_n = \pi_n^{x,\hat{r},y}$ with $x = \mu_\alpha$ and $y = \mu_\beta$. The critical step is to show that $\tilde{\sigma}$ satisfies (1.92), i.e., the expected margin of victory in the states α_1 and β_1 is larger than in the states α_2 and β_2. We show one part of (1.92), namely,

$$\lim_{n\to\infty} \max_{\omega_2 \in \{\alpha_2, \beta_2\}} |q(\tilde{\sigma}; \omega_2, \pi_n) - \frac{1}{2}| < \lim_{n\to\infty} |q(\tilde{\sigma}; \alpha_1, \pi_n) - \frac{1}{2}|. \tag{1.104}$$

The proof for the second part, the analogous statement where we replace α_1 by β_1, is verbatim with the required changes in notation. To prove (1.104), we distinguish two cases.

Case 1. $\lim_{n\to\infty} |q(\sigma^{\boldsymbol{p}}; \omega_2, \pi_n) - \frac{1}{2}| < \lim_{n\to\infty} |q(\sigma^{\boldsymbol{p}}; \alpha_1, \pi_n) - \frac{1}{2}|$.

Given (1.102), the difference is at least δ. Since almost all voters receive signal z in α_2 and β_2, the expected vote shares in α_2 and β_2 differ by much less than $\frac{\delta}{2}$ for n large enough. So, the expected margin of victory in α_1 is larger than the expected margin of victory in both α_2 and β_2 for n large enough. It follows from Claim 2 that for any $\omega_2 \in \{\alpha_2, \beta_2\}$ for which $\Pr(a|\omega_2; \pi_n, n) > 0$,

$$\lim_{n\to\infty} \frac{\Pr(\omega_2 | \text{piv}, a; \sigma^{\boldsymbol{p}}, \pi_n)}{\Pr(\alpha_1 | \text{piv}, a; \sigma^{\boldsymbol{p}}, \pi_n)} = \infty. \tag{1.105}$$

Since all voters receive a in α_1, it holds $q(\alpha_1; \tilde{\sigma}, \pi_n) = \Phi(\boldsymbol{\rho}_a(\sigma^{\boldsymbol{p}}))$. Since almost all voters receive z in α_2 and β_2 (see Figure 1.5), it holds $q(\alpha_2; \tilde{\sigma}, \pi_n) \approx \Phi(\boldsymbol{\rho}_z(\sigma^{\boldsymbol{p}}))$ and $q(\beta_2; \tilde{\sigma}, \pi_n) \approx \Phi(\boldsymbol{\rho}_z(\sigma^{\boldsymbol{p}}))$. It follows from (1.105) and Claim 7, which says that conditional on α_2 and β_2, there is nothing to be learned from the pivotal event, that, when a voter observes signal a, the inference from the signal probabilities in the states α_2 and β_2 pins down the limits of the beliefs conditional on being pivotal,

$$\begin{aligned} \lim_{n\to\infty} \Pr(\alpha | a, \text{piv}; \sigma^{\boldsymbol{p}}, \pi_n) &= \lim_{n\to\infty} \Pr(\alpha | a, \{\alpha_2, \beta_2\}; \sigma^{\boldsymbol{p}}, \pi_n) \\ &= \mu_\alpha; \end{aligned} \tag{1.106}$$

compare to (1.94). Finally, (1.104) follows from (1.106) and (1.91) together with $\Phi(\mu_\alpha) \neq \frac{1}{2}$ and $\Phi(\hat{r}) = \frac{1}{2}$. This finishes the first case.

Case 2. $\lim_{n\to\infty} |q(\sigma^{\boldsymbol{p}}; \omega_2, \pi_n) - \frac{1}{2}| > \lim_{n\to\infty} |q(\sigma^{\boldsymbol{p}}; \alpha_1, \pi_n) - \frac{1}{2}|$

Given (1.102), the difference is at least δ. Since almost all voters receive signal z in α_2 and β_2 (see Figure 1.5), the expected vote shares in α_2 and β_2 differ by much less than $\frac{\delta}{2}$ for n large enough. So, the expected margin of victory in α_1 is smaller than the expected margin of victory in both α_2 and β_2 for n large enough. It follows from Claim 2 that for $\omega_2 \in \{\alpha_2, \beta_2\}$,

$$\lim_{n\to\infty} \frac{\Pr\left(\text{piv}|\alpha_1; \sigma^{\boldsymbol{p}}, \pi_n\right)}{\Pr\left(\text{piv}|\omega_2; \sigma^{\boldsymbol{p}}, \pi_n\right)} = \infty. \tag{1.107}$$

Therefore,

$$\begin{aligned}
& \lim_{n\to\infty} \frac{\rho_a(\sigma^{\boldsymbol{p}}; \pi_n, n)}{1 - \rho_a(\sigma^{\boldsymbol{p}}; \pi_n, n)} \\
\geq\ & \lim_{n\to\infty} \frac{\Pr(\alpha) \Pr(\alpha_1|\alpha) \Pr(a|\alpha_1) \Pr\left(\text{piv}|\alpha_1; \sigma^{\boldsymbol{p}}, \pi_n\right)}{\sum_{j=1,2} \Pr(\beta) \Pr(\beta_j|\beta) \Pr(a|\beta_j) \Pr(\text{piv}|\beta_j, a; \sigma^{\boldsymbol{p}}, \pi_n)}, \\
=\ & \frac{\Pr(\alpha)}{\Pr(\beta)} \frac{(1 - \frac{r}{n^2})}{(1-r)\frac{1}{n}} \frac{1}{(1-x)\frac{1}{n^2}} \frac{\Pr\left(\text{piv}|\alpha_1; \sigma^{\boldsymbol{p}}, \pi_n\right)}{\Pr\left(\text{piv}|\beta_2; \sigma^{\boldsymbol{p}}, \pi_n\right)} \\
=\ & \infty,
\end{aligned} \tag{1.108}$$

where the equality on the third line follows since the probability of signal a is zero in β_1 and where we used (1.107) for the equality on the last line.
We will show now that (1.108) implies (1.104): to see why, recall that for n large enough, $q(\alpha_2; \tilde{\sigma}, \pi_n) \approx \Phi(\boldsymbol{\rho}_z(\sigma^{\boldsymbol{p}}; \pi_n, n))$ and $q(\beta_2; \tilde{\sigma}, \pi_n) \approx \Phi(\boldsymbol{\rho}_z(\sigma^{\boldsymbol{p}}; \pi_n, n))$ since almost all voters receive z in α_2 and β_2. Also, $q(\alpha_1; \tilde{\sigma}, \pi_n) = \Phi(\boldsymbol{\rho}_a(\sigma^{\boldsymbol{p}}; \pi_n, n))$ since all voters receive a in α_1. In addition, we have that $\rho_z(\sigma^{\boldsymbol{p}}; \pi_n, n) \approx \hat{r}$ by (1.91) and $\rho_a(\sigma^{\boldsymbol{p}}; \pi_n, n) \approx 1$ by (1.108). Finally, (1.104) follows since $\Phi(\hat{r}) = \frac{1}{2}$ and since $\Phi(1) \neq \frac{1}{2}$. This finishes the second case.

Now, we finish the proof of Claim 10. Since we just showed that, given $\tilde{\sigma} = \sigma^{\boldsymbol{\rho}(\sigma^{\boldsymbol{p}}; \pi_n, n)}$, the expected margin of victory in α_1 and β_1 is larger than in α_2 and β_2, it follows from Claim 8 that

$$\lim_{n\to\infty} \frac{\Pr\left(\{\alpha_2, \beta_2\}|\text{piv}, s; \tilde{\sigma}, \pi_n, n\right)}{\Pr\left(\{\alpha_1, \beta_1\}|\text{piv}, s; \tilde{\sigma}, \pi_n, n\right)} = \infty \tag{1.109}$$

for any $s \in \{a, b\}$. It follows from (1.109) and Claim 7, which says that conditional on α_2 and β_2, there is nothing to be learned from the pivotal event, that, given $\tilde{\sigma}$; when a voter observes signal a, the inference from the signal probabilities in the states α_2 and β_2 pins down the limits of the beliefs conditional on being pivotal, such that (1.94) and (1.95) hold for $\sigma_n = \tilde{\sigma}$. This, together with (1.91) yields Claim 10. □

1.C Persuasion of Privately Informed Voters

This section proves the following lemma that shows the "implementability" of a large set of beliefs by an appropriate choice of $(x,r,y) \in (0,1)^3$.

Lemma 3. Take any exogenous private signals π^c of the voters satisfying (1.20) and any strictly increasing Φ. There exist $0 < \lambda_\alpha < \lambda < \lambda_\beta < 1$ such that, for any $(\mu_\alpha, \mu_\beta) \in [0,1]^2$ satisfying $\mu_\alpha \notin [\lambda_\alpha, \lambda]$ and $\mu_\beta \notin [\lambda, \lambda_\beta]$, when $(x,y) \in [0,1]^2$ are given by

$$\frac{x\lambda}{x\lambda + (1-x)(1-\lambda)} = \mu_\alpha, \tag{1.110}$$

$$\frac{y\lambda}{y\lambda + (1-y)(1-\lambda)} = \mu_\beta, \tag{1.111}$$

and $r \in (0,1)$, there exists a sequence of equilibria (σ_n^*) given $\pi_n^{x,r,y}$ such that

$$\lim_{n\to\infty} \Pr\left(\alpha | \text{piv}, s_2 = a; \sigma_n^*\right) = \mu_\alpha, \tag{1.112}$$

$$\lim_{n\to\infty} \Pr\left(\alpha | \text{piv}, s_2 = z; \sigma_n^*\right) = \lambda, \tag{1.113}$$

$$\lim_{n\to\infty} \Pr\left(\alpha | \text{piv}, s_2 = b; \sigma_n^*\right) = \mu_\beta. \tag{1.114}$$

In particular, $\mu_\alpha \in \{0,1\}$ and $\mu_\beta \in \{0,1\}$ satisfy the conditions of the lemma. This implies Theorem 4.

1.C.1 Preliminaries

We provide a compact representation of equilibrium as a belief vector, similar to before in (1.13). Given any strategy σ' used by the others, the vector of posteriors conditional on piv and the additional signal $s_2 \in S_2$ is denoted as

$$\hat{\boldsymbol{\rho}}(\sigma'; \pi, n) = (\Pr(\alpha | s_2, \text{piv}; \sigma', \pi))_{s_2 \in S_2}, \tag{1.115}$$

and called the vector of *induced priors*.[28] It follows from the independence of the additional information and the exogenous information π^c that the vector of induced priors pins down the full vector of the critical beliefs: for any $s_2 \in S_2$ and any $s_1 \in \{u,d\}$,

$$\Pr(\alpha | s_1, s_2, \text{piv}; \sigma', \pi) = \frac{\hat{\rho}_{s_2}(\sigma'; \pi, n) \Pr(s_1|\alpha)}{\hat{\rho}_{s_2}(\sigma'; \pi, n) \Pr(s_1|\alpha) + (1 - \hat{\rho}_{s_2}(\sigma'; \pi, n)) \Pr(s_1|\beta)} \tag{1.116}$$

Recall that the vector of beliefs $(\Pr(\alpha | s_1, s_2, \text{piv}; \sigma', \pi))_{(s_1,s_2) \in \{u,d\} \times S_2}$ is a sufficient statistic for the unique best response to σ' for all types; see (1.11). Hence, the vector of induced priors pins down the best response for all types. Slightly abusing

28. We adopt the terminology from Bhattacharya (2013).

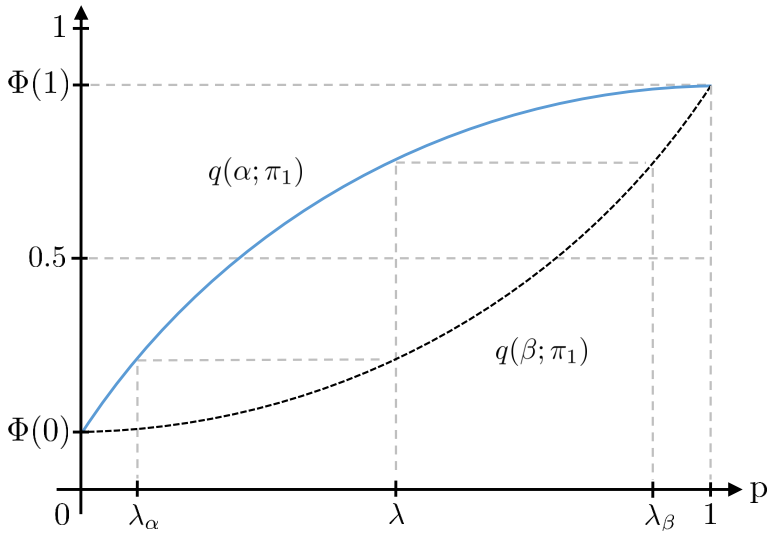


Figure 1.6. The function $\hat{q}(\alpha; p, \pi^c)$ of the implied vote share in state α and the function $\hat{q}(\beta; p, \pi^c)$ of the implied vote share in state β.

notation, for any $\boldsymbol{p} = (p_a, p_z, p_b) \in [0,1]^3$, we let $\sigma^{\boldsymbol{p}}$ be the unique strategy that is optimal given the induced prior $\boldsymbol{p}$, i.e., when a voter with signal (s_1, s_2) believes the probability of α is

$$\frac{p_{s_2} \Pr(s_1|\alpha)}{p_{s_2} \Pr(s_1|\alpha) + (1 - p_{s_2}) \Pr(s_1|\beta)}. \tag{1.117}$$

Equilibrium can be equivalently characterized by a vector of induced priors $\boldsymbol{p}^* = (p_a^*, p_z^*, p_b^*)$ such that

$$\boldsymbol{p}^* = \hat{\boldsymbol{\rho}}(\sigma^{\boldsymbol{p}^*}; \pi, n); \tag{1.118}$$

as before; see (1.13).

For any induced prior $p \in (0,1)$,

$$\hat{q}(\omega; p, \pi^c) = \sum_{s_1 \in \{u,d\}} \Pr(s_1|\omega; \pi^c) \Phi\left(\frac{p \Pr(s_1|\alpha)}{p \Pr(s_1|\alpha) + (1-p) \Pr(s_1|\beta)}\right), \tag{1.119}$$

is the probability that a voter with induced prior p draws a type t and a signal $s_1 \in S_1$ for which she votes for the outcome A in state ω. Figure 1.6 illustrates the functions $\hat{q}(\omega; p, \pi^c)$.

Since Φ is continuous and strictly increasing, it follows from (1.17) and the intermediate value theorem that there exists a unique belief λ such that the implied vote shares satisfy

$$\hat{q}(\alpha; \lambda, \pi^c) - \frac{1}{2} = \frac{1}{2} - \hat{q}(\beta; \lambda, \pi^c); \tag{1.120}$$

see Figure 1.6. Let $M = \hat{q}(\alpha; \lambda, \pi^c) - \frac{1}{2}$.

The boundaries λ_α and λ_β are such that all beliefs outside the intermediate intervals $[\lambda_\alpha, \lambda]$ and $[\lambda, \lambda_\beta]$ imply margins of victory that are larger than the ones implied by λ in *any* state $\omega \in \{\alpha, \beta\}$, i.e., larger than M. Formally, λ_α and λ_β are given by

$$q(\alpha; \lambda_\alpha, \pi^c) = q(\beta; \lambda, \pi^c), \tag{1.121}$$

$$q(\beta; \lambda_\beta, \pi^c) = q(\alpha; \lambda, \pi^c). \tag{1.122}$$

Figure 1.6 illustrates the boundaries λ_α and λ_β. For a belief $p > \lambda_\beta$,

$$\hat{q}\left(\beta; p, \pi_1\right) - \frac{1}{2} > M \tag{1.123}$$

Similarly, for $p > \lambda$,

$$\hat{q}\left(\alpha; p, \pi_1\right) - \frac{1}{2} > M \tag{1.124}$$

Note that when the exogenous information π^c of the voters becomes revealing (the signal likelihood ratios of d and u go to 0 and ∞, respectively), then

$$\lambda_\alpha \to 0, \text{ and } \lambda_\beta \to 1. \tag{1.125}$$

1.C.2 Proof of Claim 6

The Claim 6 in the main text is stated for the information structure π^r. Claim 11 below shows the analogous statement for the information structure $\pi^{x,r,y}$, noting (1.127). The same arguments imply Claim 6, and we will therefore omit its proof.

1.C.3 Voter Inference

We show that, when the sender provides additional information $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$, the induced prior after z—and thereby the margin of victory in the states α_2 and β_2—is the same across *all* equilibrium sequences and determined uniquely by the exogenous information π^c of the voters.

Claim 11. Suppose the additional information is given by $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ for some $(x,y) \in [0,1]^2$ and $r \in (0,1)$, and consider the induced sequence $(\pi_n)_{n\in\mathbb{N}}$ of independent expansions of π^c. For any equilibrium sequence (σ_n^*) given (π_n),

$$\lim_{n\to\infty} \hat{\rho}_z(\sigma_n^*, \pi_n, n) = \lambda. \tag{1.126}$$

Proof. The key idea is that, for any equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$, the election is equally close to being tied in expectation in α_2 and β_2 as $n \to \infty$.

$$\lim_{n\to\infty} q(\sigma_n^*; \alpha_2, \pi_n) - \frac{1}{2} = \lim_{n\to\infty} \frac{1}{2} - q(\sigma_n^*; \beta_2, \pi_n), \tag{1.127}$$

by arguments similar to those from the proof of the CJT; see (1.22).

Since almost all voters receive z in α_2 and β_2, the expected vote share in these states converges to the vote share implied by the induced prior after z; for $\omega_2 \in \{\alpha_2, \beta_2\}$,

$$\lim_{n\to\infty} q(\sigma_n^*; \omega_2, \pi_n) = \lim_{n\to\infty} \hat{q}(\omega; \hat{\rho}_z(\sigma_n^*; \pi_n, n), \pi^c). \tag{1.128}$$

Recall that λ is the unique induced prior such that the margins of victory are equal given the implied vote shares; see (1.120). So, (1.127) and (1.128) imply the claim, (1.126). It remains to show (1.127).

Step 1. For all n and every equilibrium σ_n^*, voters with a (z,u)-signal are more likely to vote A than voters with a (z,d)-signal when n is large enough, i.e.

$$\Phi(\rho_{z,u}(\sigma_n^*)) > \Phi(\rho_{z,d}(\sigma_n^*)). \tag{1.129}$$

This ordering follows from the likelihood ratio ordering of the signals u and d, i.e., $\frac{\Pr(u|\alpha;\pi^c)}{\Pr(u|\beta;\pi^c)} > \frac{\Pr(d|\alpha;\pi^c)}{\Pr(d|\beta;\pi^c)}$, and the independence of $\pi_n^{x,r,y}$ and π^c. Using (1.117), we have $\Pr\left(\alpha|z,u,\text{piv};\sigma_n^*,\pi_n,n\right) > \Pr\left(\alpha|z,d,\text{piv};\sigma_n^*,\pi_n,n\right)$. Now, (1.129) follows from the monotonicity of Φ.

Step 2. For all n and every equilibrium σ_n^*, the vote share of A is at most $\frac{1}{n^2}$ smaller in α_2 than in β_2,

$$q(\alpha_2; \sigma_n^*) - q(\beta_2; \sigma_n^*) \geq -\frac{1}{n^2} \tag{1.130}$$

For signals (a,b), the ordering may be the reverse of (1.129). However, in α_2 and β_2, the likelihood that a voter does not receive signal z is smaller than $\frac{1}{n^2}$. So, this follows from (1.15), given (1.20) and (1.129).

Step 3. For every equilibrium sequence (σ_n^*),

$$\lim_{n\to\infty} \hat{\rho}_z(\sigma_n^*; \pi_n, n) \notin \{0,1\}. \tag{1.131}$$

We have

$$\frac{\hat{\rho}_z(\sigma_n^*;\pi_n,n)}{1-\hat{\rho}_z(\sigma_n^*;\pi_n,n)} = \frac{\Pr(\alpha)}{\Pr(\beta)} \frac{\Pr(\alpha_2|\alpha;\pi_n)}{\Pr(\beta_2|\beta;\pi_n)} \frac{\Pr(\text{piv}|\alpha_2;\sigma_n^*,\pi_n,n)}{\Pr(\text{piv}|\beta_2;\sigma_n^*,\pi_n,n)}. \tag{1.132}$$

Suppose that $\lim_{n\to\infty} \hat{\rho}_z(\sigma_n^*;\pi_n,n) = 0$. We show that this implies

$$\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha_2;\sigma_n^*,\pi_n,n)}{\Pr(\text{piv}|\beta_2;\sigma_n^*,\pi_n,n)} \geq 1; \tag{1.133}$$

a contradiction. Since almost all voters receive z in α_2 and β_2 and since $\Phi(0) < \frac{1}{2}$, the hypothesis $\lim_{n\to\infty} \hat{\rho}_z(\sigma_n^*;\pi_n,n) = 0$ implies that

$$\lim_{n\to\infty} q(\alpha_2, \sigma_n^*) = \lim_{n\to\infty} q(\beta_2, \sigma_n^*) < \frac{1}{2}. \tag{1.134}$$

Recall that $\Phi(0) < q(\omega_j;\sigma) < \Phi(1)$ for any strategy and any substate ω_j and note that the derivative of $h(q) = q(1-q)$ is bounded below by some Lipschitz constant $L > 0$ on the compact interval $[\Phi(0), \Phi(1)]$. Hence, (1.130) implies

$$h(q(\beta_2,\sigma_n^*))(\frac{h(q(\alpha_2,\sigma_n^*))}{h(q(\beta_2,\sigma_n^*))} - 1) = h(q(\alpha_2,\sigma_n^*)) - h(q(\beta_2,\sigma_n^*)) \geq -\frac{L}{n^2}. \quad (1.135)$$

Recall that the function $h(q) = q(1-q)$ is inverse U-shaped with a peak at $q = \frac{1}{2}$ and note that it follows from (1.17) and Φ being strictly increasing that $0 < \Phi(0) < \frac{1}{2}$ and $\Phi(1) > \frac{1}{2}$. Since $\Phi(0) < q(\beta_2;\sigma_n^*) < \Phi(1)$,

$$\frac{h(q(\alpha_2,\sigma_n^*))}{h(q(\beta_2,\sigma_n^*))} \geq 1 - \frac{L}{h(q(\beta_2;\sigma_n))n^2} \geq 1 - \frac{L}{Mn^2} \quad (1.136)$$

for $M = \min\left(h(\Phi(0)), h(\Phi(1))\right)$ and all n. It follows from (1.7) that $\frac{\Pr(\text{piv}|\alpha_2;\sigma_n^*,\pi_n,n)}{\Pr(\text{piv}|\beta_2;\sigma_n^*,\pi_n,n)} \geq (1-\frac{L}{Mn^2})^n$. Now, (1.133) follows since $\lim_{n\to\infty}(1-\frac{L}{Mn^2})^n = 1$; see the analogous argument at the end of the proof of Claim 3. A similar argument excludes $\lim_{n\to\infty}\hat{\rho}_z(\sigma_n^*;\pi_n,n) = 1$ (using the analogous bound to (1.130)). This finishes the proof of the step.

Step 4. In every equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$, the limit of the vote share of A is larger in α_2 than in β_2,

$$\lim_{n\to\infty} q(\alpha_2;\sigma_n^*) > \lim_{n\to\infty} q(\beta_2;\sigma_n^*). \quad (1.137)$$

Since almost all voters receive z in α_2 and β_2, we have

$$\lim_{n\to\infty} q(\alpha_2;\sigma_n^*) = \lim_{n\to\infty} \hat{q}(\alpha;\hat{\rho}_z(\sigma_n^*,\pi_n,n)), \quad (1.138)$$

$$\lim_{n\to\infty} q(\beta_2;\sigma_n^*) = \lim_{n\to\infty} \hat{q}(\beta;\hat{\rho}_z(\sigma_n^*,\pi_n,n)). \quad (1.139)$$

From (1.131), the limits of the posteriors conditional being pivotal, the signal z and the signals $s \in \{u,d\}$ are interior, and hence, strictly ordered,

$$0 < \lim_{n\to\infty} \Pr\left(\alpha|z,d,\text{piv};\sigma_n^*,\pi_n,n\right) < \lim_{n\to\infty} \Pr\left(\alpha|z,u,\text{piv};\sigma_n^*,\pi_n,n\right) < 1. \quad (1.140)$$

Now, (1.137) follows from (1.138), (1.139), and (1.119), given (1.20), (1.140), and since Φ is strictly increasing.

We now finish the proof of Claim 11. It follows from (1.131) that voters must not become certain conditional on being pivotal and the substate being α_2 or β_2, i.e., $\lim_{n\to\infty}\Pr(\alpha|\{\alpha_2,\beta_2\},\text{piv};\sigma_n^*,\pi_n) \notin \{0,1\}$. Hence, Claim 2 requires that

$$\lim_{n\to\infty}\left|q(\alpha_2;\sigma_n^*) - \frac{1}{2}\right| = \lim_{n\to\infty}\left|q(\beta_2;\sigma_n^*) - \frac{1}{2}\right|. \quad (1.141)$$

Given the ordering of the limits of the vote shares from (1.137), the equation (1.141) implies (1.127). As noted, this completes the proof of Claim 11. □

Consider a voter who received an additional signal $s_2 \in \{a, b\}$. The following result shows that the inference from the signals is dominated by the inference from the pivotal event if the election is closer to being tied in states α_2 and β_2 than in the states α_1 and β_1. The arguments are analogous to the ones from the proof of Claims 4 and 8; we therefore omit the proof.

Claim 12. Suppose that the additional information is given by $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ for some $(x,y) \in [0,1]^2$ and $r \in (0,1)$, and consider the corresponding sequence $(\pi_n)_{n\in\mathbb{N}}$ of independent expansions of π^c. Take any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \min_{\omega_1\in\{\alpha_1,\beta_1\}} |q(\sigma_n;\omega_1,\pi_n) - \frac{1}{2}| > \lim_{n\to\infty} \max_{\omega_2\in\{\alpha_2,\beta_2\}} |q(\sigma_n;\omega_2,\pi_n) - \frac{1}{2}|, \tag{1.142}$$

then, for any $s \in \{u,d\} \times \{a,b\}$,

$$\lim_{n\to\infty} \frac{\Pr(\{\alpha_2,\beta_2\}|s,\text{piv};\sigma_n,\pi_n)}{\Pr(\{\alpha_1,\beta_1\}|s,\text{piv};\sigma_n,\pi_n)} = \infty. \tag{1.143}$$

Now, take any sequence of *equilibria* $(\sigma_n^*)_{n\in\mathbb{N}}$ that satisfies (1.142). Claim 12 implies that

$$\lim_{n\to\infty} \frac{\Pr(\alpha|a,\text{piv};\sigma_n^*,\pi_n,n)}{\Pr(\beta|a,\text{piv};\sigma_n^*,\pi_n,n)} = \lim_{n\to\infty} \frac{\Pr(\alpha_2|a,\text{piv};\sigma_n^*,\pi_n,n)}{\Pr(\beta_2|a,\text{piv};\sigma_n^*,\pi_n,n)} \tag{1.144}$$

In the following formula, we omit the dependence on σ_n^* and π_n. Using Bayes' rule,[29]

$$\begin{aligned}\lim_{n\to\infty} \frac{\Pr(\alpha_2|a,\text{piv})}{\Pr(\beta_2|a,\text{piv})} &= \lim_{n\to\infty} \frac{\Pr(\alpha)}{\Pr(\beta)} \frac{\Pr(\alpha_2|\alpha)}{\Pr(\beta_2|\beta)} \frac{\Pr(a|\alpha_2)}{\Pr(a|\beta_2)} \frac{\Pr(\text{piv}|\alpha_2)}{\Pr(\text{piv}|\beta_2)} \\ &= \lim_{n\to\infty} \frac{\Pr(\alpha|\{\alpha_2,\beta_2\},\text{piv})}{\Pr(\beta|\{\alpha_2,\beta_2\},\text{piv})} \frac{\Pr(a|\alpha_2)}{\Pr(a|\beta_2)}. \end{aligned} \tag{1.145}$$

Note that $\lim_{n\to\infty} \hat{\rho}_z(\sigma_n^*;\pi_n,n) = \lim_{n\to\infty} \Pr(\alpha|\{\alpha_2,\beta_2\},\text{piv};\sigma_n^*,\pi_n,n)$ such that Claim 11 implies

$$\lim_{n\to\infty} \Pr(\alpha|\{\alpha_2,\beta_2\},\text{piv};\sigma_n^*,\pi_n,n) = \lambda. \tag{1.146}$$

Using (1.144), (1.145), (1.146), and the definition of the information structure $\pi_n^{x,r,y}$, we conclude

$$\lim_{n\to\infty} \frac{\Pr(\alpha|a,\text{piv};\sigma_n^*,\pi_n)}{\Pr(\beta|a,\text{piv};\sigma_n^*,\pi_n)} = \frac{x}{1-x}\frac{\lambda}{1-\lambda}. \tag{1.147}$$

Similarly, for the additional signal b,

$$\lim_{n\to\infty} \frac{\Pr(\alpha|b,\text{piv};\sigma_n^*,\pi_n,n)}{\Pr(\beta|b,\text{piv};\sigma_n^*,\pi_n,n)} = \frac{y}{1-y}\frac{\lambda}{1-\lambda}. \tag{1.148}$$

29. As before, if $x = 1$, that is if $\Pr(a|\beta_2) = 0$, using the convention $\frac{1}{0} = \infty$, the following equality holds in the extended reals.

1.C.4 Fixed Point Argument

In this section, we prove Lemma 3, using the observations from the preceding section. Let us consider some belief $\mu_\alpha \notin [\lambda_\alpha, \lambda]$ and some belief $\mu_\beta \notin [\lambda, \lambda_\beta]$ with λ, λ_α, and λ_β given by (1.120), (1.121) and (1.122).
Recall from Section 1.C.1 that equilibrium can be equivalently characterized by a vector of beliefs $\boldsymbol{p}^* = (p_a^*, p_z^*, p_b^*)$ such that $\boldsymbol{p}^* = \hat{\boldsymbol{\rho}}(\sigma^{\boldsymbol{p}^*}; \pi, n)$; see (1.118). Now, take any $\delta > 0$ and let

$$\mathrm{B}_\delta = \left\{ \boldsymbol{p} \in [0,1]^3 \,|\, |\boldsymbol{p} - (\mu_\alpha, \lambda, \mu_\beta)| \leq \delta \right\}.$$

Take any $\boldsymbol{p} \in \mathrm{B}_\delta$ and the corresponding strategy $\sigma^{\boldsymbol{p}}$. We define a constrained best-response function as its "truncation" to B_δ:

$$\hat{\rho}_a^{tr}(\sigma^{\boldsymbol{p}}) \quad = \quad \begin{cases} \mu_\alpha - \delta & \text{if} \quad \hat{\rho}_a(\sigma^{\boldsymbol{p}}) < \mu_\alpha - \delta, \\ \mu_\alpha + \delta & \text{if} \quad \hat{\rho}_a(\sigma^{\boldsymbol{p}}) > \mu_\alpha + \delta, \\ \hat{\rho}_a(\sigma^{\boldsymbol{p}}) & \text{else.} \end{cases} \tag{1.149}$$

The components $\hat{\rho}_z^{tr}$ and $\hat{\rho}_b^{tr}$ are defined in the analogous way. The function $\hat{\boldsymbol{\rho}}^{tr}(\sigma^{\boldsymbol{p}})$ is continuous in $\boldsymbol{p}$ such that Kakutani's theorem implies that $\hat{\boldsymbol{\rho}}^{tr}(\sigma^{\boldsymbol{p}})$ has a fixed point $\boldsymbol{p}^* \in \mathrm{B}_\delta$.

Any fixed point $\boldsymbol{p}^*$ of $\hat{\boldsymbol{\rho}}^{tr}$ is shown to be in the interior of B_δ when n is large enough and δ is small enough, i.e., $\hat{\boldsymbol{\rho}}^{tr}(\sigma^{\boldsymbol{p}^*}) = \hat{\boldsymbol{\rho}}(\sigma^{\boldsymbol{p}^*})$:

Claim 13. Consider any $\mu_\alpha \notin [\lambda_\alpha, \lambda]$ and any $\mu_\beta \notin [\lambda, \lambda_\beta]$. Consider the sequence of independent expansions $(\pi_n)_{n\in\mathbb{N}}$ of π^c with additional information $(\pi_n^{x,r,y})_{n\in\mathbb{N}}$ where $\mu_\alpha = \frac{x\lambda}{x\lambda+(1-x)(1-\lambda)}$ and $\mu_\beta = \frac{y\lambda}{y\lambda+(1-y)(1-\lambda)}$ and $r \in (0,1)$.

For any $\delta > 0$ small enough, there exists $n(\delta) \in \mathbb{N}$ such that for all $n \geq n(\delta)$, any fixed point of $\hat{\boldsymbol{\rho}}^{tr}$ is in the interior of B_δ.

Proof. Pick some $\boldsymbol{p}$ for which p_z is on the boundary. We show $\boldsymbol{p}$ cannot be a fixed point for n large enough and δ small enough. First, suppose $p_z = \lambda - \delta$. Then, given σ and as $n \to \infty$, the margin of victory in α_2 is strictly smaller than the margin of victory in β_2, given the definition of λ; see (1.120). Hence, Claim 2 implies that $\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha_2;\sigma^{\boldsymbol{p}},\pi_n^{x,r,y},n)}{\Pr(\text{piv}|\beta_2,\sigma^{\boldsymbol{p}},\pi_n^{x,r,y},n)} = \infty$. This implies, $\lim_{n\to\infty} \hat{\rho}_z(\sigma^{\boldsymbol{p}}; \pi_n^{x,r,y}, n) = 1$. For any n large enough this contradicts $p_z = \lambda - \delta$ and so $\boldsymbol{p}$ is not a fixed point of $\hat{\boldsymbol{\rho}}^{tr}(\sigma^{\boldsymbol{p}})$. In the same way we can exclude that $p_z = \lambda + \delta$ for any n large enough. In general, the same argument implies that, for n large enough, for any fixed point $\boldsymbol{p}^*$,

$$\hat{\rho}_z(\sigma^{\boldsymbol{p}^*}) \approx \lambda. \tag{1.150}$$

Given the assumptions on μ_α, μ_β, we can choose $\delta > 0$ small enough such that, for any $\boldsymbol{p} \in \mathrm{B}_\delta$ and the corresponding behavior $\sigma^{\boldsymbol{p}}$, the expected margins of victory in the states α_2 and β_2 are strictly smaller than the expected margins of victory in the

states α_1 and β_1, i.e., $\sigma^{\boldsymbol{p}}$ satisfies (1.142). Therefore, it follows from Claim 12 and (1.150) that (1.147) and (1.148) hold; hence, given the definition of $\hat{\boldsymbol{\rho}}$

$$\hat{\rho}_a(\sigma^{\boldsymbol{p}^*}) \approx \mu_\alpha, \tag{1.151}$$

$$\hat{\rho}_b(\sigma^{\boldsymbol{p}^*}) \approx \mu_\beta. \tag{1.152}$$

We conclude that any fixed point $\boldsymbol{p}^*$ of $\hat{\boldsymbol{\rho}}^{tr}$ is interior when δ is small enough and n is large enough. □

Now, we finish the proof of Lemma 3. Note that the strategy $\sigma^{\boldsymbol{p}^*}$ corresponding to any interior fixed point $\boldsymbol{p}^*$ of $\hat{\boldsymbol{\rho}}^{tr}$ is an equilibrium. Therefore, Claim 13 implies the existence of a sequence of equilibria $(\sigma_n^*)_{n\in\mathbb{N}}$ for which (1.150), (1.151), and (1.152) hold. This finishes the proof of Lemma 3.

References

Acharya, Avidit. 2016. "Information aggregation failure in a model of social mobility." *Games and Economic Behavior* 100: 257–272. [36]

Acharya, Avidit, and Adam Meirowitz. 2017. "Sincere voting in large elections." *Games and Economic Behavior* 101: 121–131. [36]

Ahn, David S, and Santiago Oliveros. 2012. "Combinatorial voting." *Econometrica* 80 (1): 89–141. [12]

Ali, S Nageeb, Maximilian Mihm, and Lucas Siga. 2018. "Adverse Selection in Distributive Politics." Working paper. [36]

Alonso, Ricardo, and Odilon Câmara. 2016. "Persuading voters." *American Economic Review* 106 (11): 3590–3605. [7, 29, 34]

Bardhi, Arjada, and Yingni Guo. 2018. "Modes of persuasion toward unanimous consent." *Theoretical Economics* 13 (3): 1111–1149. [7, 22, 33, 34]

Barelli, Paulo, Sourav Bhattacharya, and Lucas Siga. 2019. "Full information equivalence in large elections." Working Paper. [36]

Bergemann, Dirk, Benjamin A Brooks, and Stephen Morris. 2016. "Informationally robust optimal auction design." Working Paper. [37]

Bergemann, Dirk, and Stephen Morris. 2016. "Bayes correlated equilibrium and the comparison of information structures in games." *Theoretical Economics* 11 (2): 487–522. [34]

Bergemann, Dirk, and Stephen Morris. 2019. "Information design: A unified perspective." *Journal of Economic Literature* 57 (1): 44–95. [34]

Bhattacharya, Sourav. 2013. "Preference monotonicity and information aggregation in elections." *Econometrica* 81 (3): 1229–1247. [12, 15, 36, 49]

Bhattacharya, Sourav. 2018. "Condorcet Jury theorem in a spatial model of elections." Working paper. [36]

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang. 2019. "Pivotal persuasion." *Journal of Economic Theory* 180: 178–202. [7, 22, 33, 34]

De Clippel, Geoffroy, Rene Saran, and Roberto Serrano. 2019. "Level-Mechanism Design." *Review of Economic Studies* 86 (3): 1207–1227. [27]

Du, Songzi. 2018. "Robust mechanisms under common valuation." *Econometrica* 86 (5): 1569–1588. [37]

Ekmekci, Mehmet, and Stephan Lauermann. 2019. "Manipulated Electorates and Information Aggregation." *Review of Economic Studies*, Forthcoming. [36]

Feddersen, Timothy, and Wolfgang Pesendorfer. 1997. "Voting behavior and information aggregation in elections with private information." *Econometrica:* 1029–1058. [5–7, 14, 15, 35–37]

Feng, Tangren, and Qinggong Wu. 2019. "Getting Information from Your Enemies." Working Paper. [35]

Gerardi, Dino, Richard McLean, and Andrew Postlewaite. 2009. "Aggregation of expert opinions." *Games and Economic Behavior* 65 (2): 339–371. [35]

Guo, Yingni, and Eran Shmaya. 2019. "The interval structure of optimal disclosure." *Econometrica* 87 (2): 653–675. [35]

Heese, Carl, and Stephan Lauermann. 2019. "Persuasion and Information Aggregation in Large Elections." CRC TR 224 Discussion Paper Series. University of Bonn, and Univer-

sity of Mannheim, Germany. https://ideas.repec.org/p/bon/boncrc/crctr224_2019_128.html. [27, 32, 33, 35, 36]

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review* 101 (6): 2590–2615. [5]

Kerman, Toygar, P Jean-Jacques Herings, and Dominik Karos. 2019. "Persuading Voters With Private Communication Strategies." Working Paper. [35]

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. 2017. "Persuasion of a privately informed receiver." *Econometrica* 85 (6): 1949–1964. [35]

Kosterina, Svetlana. 2019. "Information Structures and Information Aggregation in Threshold Equilibria in Elections." Working Paper. [36]

Levy, Gilat, Inés Moreno de Barreda, and Ronny Razin. 2018. "Persuasion with correlation neglect: media power via correlation of news content." Working Paper. [35]

Mandler, Michael. 2012. "The fragility of information aggregation in large elections." *Games and Economic Behavior* 74 (1): 257–268. [36]

Mathevet, Laurent, Jacopo Perego, and Ina Taneva. 2017. "On information design in games." Working Paper. [34]

McLennan, Andrew. 1998. "Consequences of the Condorcet jury theorem for beneficial information aggregation by rational agents." *American political science review* 92 (2): 413–418. [28]

Razin, Ronny. 2003. "Signaling and election motivations in a voting model with common values and responsive candidates." *Econometrica* 71 (4): 1083–1119. [36]

Salcedo, Bruno. 2019. "Persuading part of an audience." Working Paper. [35]

Schipper, Burkhard, and Hee Yeul Woo. 2019. "Political awareness and microtargeting of voters in electoral competition." *Quarterly Journal of Political Science* 14: 41–88. [35]

Schnakenberg, Keith E. 2015. "Expert advice to a voting body." *Journal of Economic Theory* 160: 102–113. [35]

Wang, Yun. 2013. "Bayesian persuasion with multiple receivers." Working Paper. [7, 34]

Weinstein, Jonathan, and Muhamet Yildiz. 2007. "A structure theorem for rationalizability with application to robust predictions of refinements." *Econometrica* 75 (2): 365–400. [8]

Yamashita, Takuro et al. 2016. "Revenue Guarantee in Auction with a (Correlated) Common Prior and Additional Information." Working Paper. [37]

Chapter 2

Voter Attention and Distributive Politics

2.1 Introduction

This paper studies how endogenous attention to politics affects social welfare and its distribution. It is guided by the empirical observation that voters that care more about a political issue will acquire more information about it.[1]

I propose a model of an election over a distributive reform with uncertain consequences.Examples of such reforms are numerous: a trade reform opens new markets for exporting firms but threatens the prospects in other sectors;a public health policy reform makes certain treatments more accessible to some citizens, while implying price increases for a range of pharmaceuticals needed by others; and a new education reform benefits some children but affects others negatively. In all these examples, some voters are *ex-ante* uninformed about the consequences of the reform, e. g. which sectors gain from a trade reform, or if their child benefits from education reform. However, they hold private information about their exposure to the proposed reform, that is, about the magnitude of their preference intensities: older people care more about healthcare issues, while changes in education policy are more relevant to citizens with children (Iyengar, Hahn, Krosnick, and Walker, 2008).

Are referenda and elections efficient mechanisms of collective choice in such situations? This paper considers a modified version of the canonical setting by Feddersen and Pesendorfer (1997). Relative to Feddersen and Pesendorfer (1997), the voters' information about the policies is endogenous *and* the setting allows that the voters have conflicting interests (distributive politics). There are two possible policies: a reform and a status quo. Voters' preferences over policies are heterogeneous

1. This is known as the "issue publics hypothesis" (Converse, 1964). See e.g. Krosnick (1990) and Henderson (2014), and Carpini and Keeter (1996) for an overview about the American public's factual knowledge about politics.

and depend on an unknown, binary state in a general way (some voters may prefer the reform only in the first state, others may prefer the reform only in the second state, while some "partisans" may prefer one of the policies independently of the state). The preferences are each voter's private information. Besides, all voters can receive information about the state in the form of a noisy signal, and each voter freely chooses the precision of her private signal. More precise information is more costly. Upon receiving their signals, all citizens vote simultaneously. The election determines the outcome by simple majority rule. Feddersen and Pesendorfer (1997) have shown that when voters receive conditionally i.i.d. signals of some exogenous quality and their preferences are "monotone" in all equilibria of large elections the outcome preferred by the median voter is elected state-by-state.[2] In many situations where voters have conflicting interests this is not the first-best outcome: for example, when 51% of the citizens marginally benefit from a reform, while the other 49% are severely impacted by it.

In our setting, elections either lead to the full-information outcome, but otherwise lead to outcomes that are only preferred by a minority ex-post (Theorem 2). This is the case when a minority of the voters is more severely affected by the reform. As a consequence, the minority will be better informed. Importantly, the more information the minority voters acquire, the more they correlate their vote with the unknown state of the world, thereby pushing the outcome in their favorite direction in each state. When voters of the minority group acquire substantially more information than others, they coordinate well on voting for their preferred policy, and this policy will indeed be elected in each state.

We provide the result that election outcomes are as if the decision weight of a citizen is proportional to how informed she is, provided the cost of voters to acquire political information are not "too high". This has important implications: first, when information cost are extremely low, all voters are relatively well informed, the implicit decision weights of citizens are approximately equal, and election outcomes almost always lead to majority-preferred outcomes. Second, uninformed citizens have no voting power, similar to voters that abstain due to voting being costly. Third, elections may be susceptible to targeted informational interventions of third-parties, which we will discuss in Section 2.6.

The main result characterizes which group of voters sharing a common interest will win the election. For this, we aggregate the decision weights of the citizens to describe the *power* of voter groups with common interests. A group's power increases in its size and the group's welfare at stake. The main result shows that the group with the larger power wins the election in each state. Under an independence condition,this yields sharp welfare predictions: elections lead to outcomes that maximize a weighted welfare rule. The weight of a voter is higher when her utilities are

2. The preference distribution of the voters is "monotone" if a higher belief in the first state entails that more voters prefer the reform.

higher, but less so, when information is more cheap. For example, for intermediate cost, each citizen's information and weight turn out to be proportional to her utility. Then, elections lead to utilitarian outcomes.[3]

The main result describes the properties of limit equilibria with state-dependent election outcomes in large electorates. Thereby, we show, in particular, existence of such *informative* limit equilibria. This is economically surprising since voters of a large electorate face a severe free-rider problem when acquiring private information is costly, much similar to the reasoning in Downs' paradox of voting (Downs, 1957). The existence of informative limit equilibria relies on the observation that information acquisition in elections can be complementary, which we discuss in Section 2.4.5.4. This complementarity also drives an equilibrium multiplicity. Citizens may coordinate on paying much or very few attention to politics (Theorem 4).

In Section 2.6, we provide several extensions: first, we discuss the role of polarization of utilities within voter groups, and show that a more polarized group has a smaller electoral power and sufficiently much polarization, ceteris paribus, will lead the group to lose the election (Theorem 6). Second, we provide an extension where the cost of information of voters is heterogeneous, capturing that citizens have different abilities to access and interpret political information. Third, we discuss the potential of manipulative information provision by third-parties and its effectiveness.

In Section 2.7, we discuss the paper's contribution to the literature on voting cost and vote buying, especially Krishna and Morgan (2011) and Lalley and Weyl (2018). We also discuss the contribution to the literature on distributive politics, especially Fernandez and Rodrik (1991), and to the literature on information aggregation in elections: both modifications relative to Feddersen and Pesendorfer (1997) that are made in this paper have been studied before, but not together: Martinelli (2006) has studied a variant with endogenous information, and shown that the median voter theorem also holds, but only if voters can acquire relevant political information at a cost that is "not too high," thereby establishing the first existence result for informative equilibrium sequences.[4] Bhattacharya (2013a) has shown that the median voter result generalizes to settings with conflicting interests.[5] Importantly, his model does not allow to study the role of the intensity of preferences since the result is invariant

3. I also show that aggregate cost of the voters converge to 0 as the electorate grows large such that the equilibrium sequences with utilitarian outcomes imply first-best results, even when taking into account the cost of voters, see Lemma 12.

4. Formally, what matters for the result is how fast cost goes to zero when a voter chooses an arbitrarily uninformative signal. Basically, the critical condition is that elasticity of the cost function at the precision of the uninformative signal is large enough. The same condition is necessary in this paper for the existence of limit equilibria with non-trivial state-dependent outcomes, as the electorate grows large.

5. Bhattacharya (2013a) also shows that the result breaks down when preferences are non-monotone; in particular, even minimal non-monotonicities turn around welfare predictions.

to scaling the intensities of specific groups of voters. The paper also relates to a literature studying the interaction of limited attention of voters and the policy choices of political platforms. Matějka and Tabellini (2017) study this question in a probabilistic voting model.[6] There, citizens who pay attention are more responsive to policy changes, and as a consequence, political candidates offer policies catered to more attentive citizens. What differs is that in their work, endogenous attention distorts equilibrium policies away from first-best policies; in other words, the welfare implication is in the opposite direction relative to this paper. Second, the mechanism how attention affects policy outcomes is distinct from this paper, where information implicitly allows voter groups to coordinate more strongly, thereby enhancing their electoral power.

A Two-Type Example

The following extreme setting shows how a minority can overcome the dominance of a majority by correlating their vote more strongly with the state than the majority. Thereby, we illustrate how utilitarian outcomes can be elected, even when a majority of the voters do not prefer the outcome *ex-post*.[7] There are $2n+1$ voters. With probability $1 > \lambda > \frac{1}{2}$, a voter is *aligned* and prefers the reform A over the status quo B only in α and B over A in β. Otherwise, a voter is *contrarian* and prefers A in β and B in α.[8] An aligned voter gets a small utility of $\epsilon > 0$ when her preferred policy is adopted, while a contrarian voter gets a utility of 1 when her preferred policy is adopted. Each voter can either get a private, perfect signal about the state at a given cost $c > 0$ or an uninformative signal at no cost. The common prior about the state is uniform, i.e., $\Pr(\alpha) = \frac{1}{2}$. Let ϵ is sufficiently small such that $\epsilon\lambda < (1-\lambda)$; hence, in order to maximize utilitarian welfare, the election should choose the contrarians' prefered policy.

Consider three scenarios: zero, intermediate and high cost. When cost is zero, i.e. $c = 0$, all voters become perfectly informed about the state and the outcome preferred by the median voter is elected in each state. When the cost is very high, e. g. $c > 1$, nobody gets informed, and the policy elected is independent of the state. Now, suppose that only the contrarians receive the perfect signal, and vote for their prefered outcome in each state; the aligned have no information about the state and vote for each policy with the same probability, i.e. $50-50$. Then, in each state, the outcome preferred by the contrarians is elected as the electorate grows large. We claim that this behaviour is an equilibrium for an intermediate range of cost c. The

6. See also Hu and Li (2018).

7. We do not discuss the welfare effects of the cost for the example since it turns out that the aggregate cost is arbitrarily small in the equilibria of the main model when the electorate is large, $n \to \infty$ (see Section 2.6.4).

8. The terminology used to label the voter types carries no economic meaning whatsoever but only relates to the notation. Aligned voters prefer the outcome that is "aligned" with the state.

relevant observation is that the value of information is higher for the contrarians since $\epsilon < 1$. As a consequence, there are intermediate levels of cost that exceed the value of information for the aligned, but not for the contrarians.[9]

2.2 Model

There are $2n+1$ voters (or citizens), two policies A and B, and two states of the world $\omega \in \{\alpha, \beta\} = \Omega$. The prior probability of α is $\Pr(\alpha) \in (0,1)$.

Voters have heterogeneous and state-dependent preferences. A voter's preference is described by a type $t = (t_\alpha, t_\beta)$, where $t_\omega \in [-1,1]$ is the utility of A in ω. The utility of B is normalized to zero, so that t_ω is the difference between the utilities of A and B in ω. The types are identically distributed across voters according to a cumulative distribution function $H : [-1,1]^2 \to [0,1]$ that has a continuous density h. A voter's type is her private information. Each voter privately observes a binary signal $s \in \{a,b\}$ about the state. The joint distribution of the type and the signal of a voter is independent of the distribution of the signals and the types of the other voters conditional on the state.

The voting game is as follows. First, nature draws the state and the profile of types $\boldsymbol{t}$ according to H. Second, after observing her type, each voter chooses a precision $x(t) \in [0, \frac{1}{2}]$ of her signal, that is $\frac{1}{2} + x(t) = \Pr(a|\alpha) = \Pr(b|\beta)$. Then, private signals realize. After observing her private signal, each voter simultaneously submits a vote for A or B. Finally, the submitted votes are counted and the majority outcome is chosen.

There is a strictly increasing, strictly convex, and twice continuously differentiable cost function $c : [0, \frac{1}{2}] \to \mathbb{R}_+$ and when choosing precision x, the voter bears a cost $c(x)$ where $c(0) = 0$. There is $d > 1$ such that[10]

$$\lim_{x\to\infty} \frac{c'(x)}{x^{d-1}} \in \mathbb{R}. \tag{2.1}$$

A strategy $\sigma = (x, \mu)$ of a voter consists of a function $x : [-1,1]^2 \to [0, \frac{1}{2}]$ mapping types to signal precisions and of a function $\mu : [-1,1]^2 \times \{a,b\} \to [0,1]$ mapping types and signals to probabilities to vote A, i.e., $\mu(t,s)$ is the probability that a voter

9. For completeness, note that without the private signal, a citizen is indifferent between voting for either of the policies. First, recall that the prior is $\Pr(\alpha) = \frac{1}{2}$. Second, the citizens do not infer anything about the state from conditioning on being pivotal for the election outcome. This is because the event in which the citizen's vote affects the outcome is equally likely in each state in the candidate equilibrium since in β the reform wins with the same margin of $\left[\lambda\frac{1}{2} + (1-\lambda)\right] - \frac{1}{2} = \frac{1}{2}(1-\lambda)$ and in α the reform loses with a margin $\frac{1}{2}(1-\lambda)$ in expectation.

10. It will be a direct insight from the preliminary results in the next section that without the condition $d > 1$, no voter acquires any information in equilibrium when n is sufficiently large; see (2.18).

of type t with signal s votes for A. I consider only non-degenerate strategies.[11] I analyze the Bayes-Nash equilibria of the Bayesian game of voters in symmetric strategies, henceforth called *equilibria*.

2.2.1 Preferences

Figure 2.1 shows the area of possible preference types. Voters having types t in the north-east quadrant prefer A for all beliefs and voters having types t in the south-west quadrant always prefer B (*partisans*). Voters having types t in the south-east quadrant prefer A in state α and B in β (*aligned voters*), and voters having types t in the north-west quadrant prefer B in state α and A in β (*contrarian voters*). To simplify the exposition, in the rest of the paper, we only consider strategies σ where the partisans use the (weakly) dominant strategy to vote for their preferred policy.[12]

Aggregate Preferences. A central object of the analysis is the *aggregate preference function*

$$\Phi(p) = \Pr_H(\{t : p \cdot t_\alpha + (1-p) \cdot t_\beta \geq 0\}), \tag{2.2}$$

which maps a belief $p \in [0,1]$ about the state being α to the probability that a random type t prefers A given p. Figure 2.1 illustrates Φ: the (colored) line corresponds to the set of types $t = (t_\alpha, t_\beta)$ that are indifferent between policy A and policy B when holding the belief p. Voters having types to the north-east prefer A given p (shaded area); these types have mass $\Phi(p)$. The indifference set has a slope of $\frac{-p}{1-p}$ and an increase in p corresponds to a clockwise rotation of it. Given that H has a continuous density, Φ is continuously differentiable in p.

I assume that

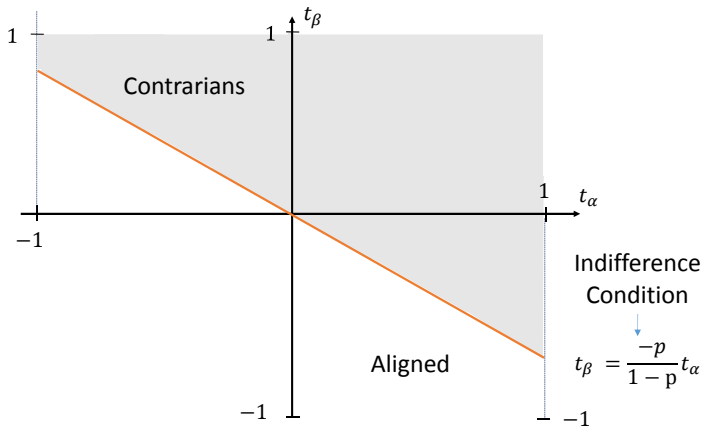
$$\Phi(0) < \frac{1}{2}, \text{ and } \Phi(1) > \frac{1}{2} \tag{2.3}$$

such that the median-voter preferred outcome is A in α and B in β. In particular, this excludes the cases when there is a majority of partisans for one policy in expectation. I also make the genericity assumption that Φ is not constant on any open interval.[13] Henceforth, I will call distributions H that have a continuous density and satisfy (2.3) simply *preference distributions*. The set of the aligned types is denoted $L = \{t : t_\alpha >$

11. A strategy σ is *degenerate* if $\mu(t,s) = 1$ for all (t,s) or if $\mu(t,s) = 0$ for all (s,t). When all voters follow the same degenerate strategy and there are at least three voters, if one voter deviates to any other strategy, then the outcome is the same. Therefore, the degenerate strategies with $x = 0$ are trivial equilibria.

12. In fact, for any non-degenerate strategy, I show that the likelihood that a given voter is pivotal for the election outcome is non-zero (see Section 2.3.1) such that voting for the preferred policy while not acquiring any information is the unique strict best response for all partisans.

13. This assumption is known from the literature, see Bhattacharya (2013b).



Notes: For any given belief $p = \Pr(\alpha) \in (0,1)$, the set of types t with a threshold of doubt $y(t) = p$ is given by $t_\beta = \frac{-p}{1-p} t_\alpha$. Voter types north-east of the indifference line (shaded area) prefer A given p.

Figure 2.1. The preference types

$0, t_\beta < 0\}$ and the set of the contrarian types is denoted $C = \{t : t_\alpha < 0, t_\beta > 0\}$ and $g \in \{L, C\}$ is the generic symbol for a voter group, aligned or contrarians.

Threshold of Doubt and Preference Intensity It is useful to view types as information about, first, the relative preference intensities across states,

$$y(t) = \frac{-t_\beta}{t_\alpha - t_\beta}, \tag{2.4}$$

and, second, the *total intensity*,

$$k(t) = t_\alpha - t_\beta. \tag{2.5}$$

For any aligned type $t, y(t)$ and $k(t)$ together uniquely pin down t.[14] Similarly, for any contrarian type t, $y(t)$ and $k(t)$ together uniquely pin down t. Recall that a strategy describes a voting choice and an information choice for each type. Section 2.3 shows that the *threshold of doubt* $y(t)$ determines the voting choice of (non-partisan) types, and the total intensity determines the information choice.

2.3 Citizens' Votes and Information

2.3.1 Threshold of Doubt Pins Down Vote

Take any strategy $\sigma = (x, \mu)$ of the voters. The probability that a voter of random type votes for A in state $\omega \in \{\alpha, \beta\}$ is denoted $q(\omega; \sigma)$. A simple calculation shows

14. For $t \in L$, $y(t)k(t) = t_\beta$, and $(1 - y(t))k(t) = t_\alpha$.

that

$$q(\alpha;\sigma) = \int_{t\in[-1,1]^2} (\frac{1}{2}+x(t))\mu(t,a)+(\frac{1}{2}-x(t))\mu(t,b)dHt, \quad (2.6)$$

and

$$q(\beta;\sigma) = \int_{t\in[-1,1]^2} (\frac{1}{2}-x(t))\mu(t,a)+(\frac{1}{2}+x(t))\mu(t,b)dHt. \quad (2.7)$$

I also refer to $q(\omega;\sigma)$ as the *(expected) vote share* of A in ω.

Pivotal Voting. Take a single citizen, and fix a strategy σ' of the other voters. The given citizen's vote determines the outcome only in the event when the votes of the other citizens tie, denoted piv. Thus, a strategy is optimal if and only if it is optimal conditional on the pivotal event piv. The probability that the votes of the other citizens tie in ω is

$$\Pr(\text{piv}|\omega;\sigma',n) = \binom{2n}{n}\left(q(\omega;\sigma')\right)^n\left(1-q(\omega';\sigma)\right)^n. \quad (2.8)$$

since conditional on the state, the type and the signal of a voter is independent of the types and the signals of the other voters. For any type t of the given citizen, and given the precision choice $x(t)$, let $\Pr(\alpha|s,\text{piv};\sigma',n)$ be the posterior probability of α conditional on having received the private signal s and conditional on *being pivotal* when the other voters use σ'. We conclude that, μ is part of a best response $\sigma=(x,\mu)$ if and only if for all $t=(t_\alpha,t_\beta)$ and for the signal precision $x(t)$,

$$\Pr(\alpha|s,\text{piv};\sigma',n)\cdot t_\alpha+(1-\Pr(\alpha|s,\text{piv};\sigma',n))\cdot t_\beta>0\Rightarrow\mu(s,t)=1, \quad (2.9)$$
$$\Pr(\alpha|s,\text{piv};\sigma',n)\cdot t_\alpha+(1-\Pr(\alpha|s,\text{piv};\sigma',n))\cdot t_\beta<0\Rightarrow\mu(s,t)=0, \quad (2.10)$$

that is, a voter supports A if the expected value of A conditional on being pivotal and s is strictly positive and otherwise supports B. Note that for each aligned type $t\in L$, (1.3) and (1.4) are equivalent to

$$\Pr(\alpha|s,\text{piv};\sigma',n)>y(t)\Rightarrow\mu(t,s)=1, \quad (2.11)$$
$$\Pr(\alpha|s,\text{piv};\sigma',n)<y(t)\Rightarrow\mu(t,s)=0; \quad (2.12)$$

and for all contrarian types $t\in C$, (1.3) and (1.4) are equivalent to

$$\Pr(\alpha|s,\text{piv};\sigma',\sigma,n)>y(t)\Rightarrow\mu(t,s)=0, \quad (2.13)$$
$$\Pr(\alpha|s,\text{piv};\sigma',\sigma,n)<y(t)\Rightarrow\mu(t,s)=1, \quad (2.14)$$

We see that $y(t)$ is the unique belief that a makes a voter of type t indifferent, thereby qualifying the name *threshold of doubt*.

2.3.2 Preference Intensity Pins Down Information Level

What is the marginal value of information to a citizen? Take an aligned voter, and fix the likelihood $x(t) > 0$ of her receiving a correct signal about the state. At the end of this section, we establish that she votes A after a and B after b (Lemma 1), that is, she votes for her preferred policy in each state whenever receiving a "correct signal" . When she is not pivotal, the policy elected is independent of her private precision $x(t)$. In the pivotal event, using Lemma 1, her expected utility from the elected policy is

$$\Pr(\text{piv}|\sigma', n)\Pr(\alpha|\text{piv};\sigma)(\frac{1}{2} + x(t))t_\alpha \tag{2.15}$$

in state α, and

$$\Pr(\text{piv}|\sigma', n)\Pr(\beta|\text{piv};\sigma)(\frac{1}{2} - x(t))t_\beta \tag{2.16}$$

in state β, where we used that the utility from B is normalized to zero.[15] Therefore, the marginal benefit of a higher precision $x(t)$ is

$$\begin{aligned} & MB(x(t);\sigma', n) \\ = \;& \Pr(\text{piv}|\sigma', n)(\Pr(\alpha|\text{piv};\sigma)t_\alpha - \Pr(\beta|\text{piv};\sigma)t_\beta) \\ = \;& \Pr(\text{piv}|\sigma', n)k(t)c_1(y(t)) \end{aligned} \tag{2.17}$$

for $c_1(y(t)) = \Pr(\alpha|\text{piv};\sigma)(1 - y(t)) + \Pr(\beta|\text{piv};\sigma)y(t)$, where we used that $t_\alpha = k(t)(1 - y(t))$ and $t_\beta = k(t)y(t)$ for the last equation. We see that the total intensity $k(t)$ is decisive. Finally, for any type t for which it is optimal to acquire some information, $x(t) > 0$, the precision is pinned down by equating marginal benefits and marginal cost,

$$c'(x(t)) = MB(x(t);\sigma', n) \tag{2.18}$$

It follows from the strict convexity of c, that for any t, there is a unique solution to (2.18), denoted $x^*(t;\sigma', n)$. Moreover, $x^*(t;\sigma', n)$ is continuously differentiable by an application of the implicit function theorem, and, given (2.1),

$$x^*(t;\sigma, n) \approx MB(x(t);\sigma', n)^{\frac{1}{d-1}}. \tag{2.19}$$

15. Similarly, in the pivotal event, a contrarian's expected utility when choosing $x(t)$ is

$$\Pr(\text{piv};\sigma', n)\Pr(\alpha|\text{piv};\sigma)(\frac{1}{2} - x(t))t_\alpha$$

in state α, and

$$\Pr(\text{piv};\sigma', n)\Pr(\beta|\text{piv};\sigma)(\frac{1}{2} + x(t))t_\beta$$

in state β.

Lemma 1. Take any strategy σ'. The function μ is part of a best response $\sigma = (x, \mu)$ if and only if

$$\forall t \in L : x(t) > 0 \Rightarrow \mu(t, a) = 1 \text{ and } \mu(t, b) = 0, \tag{2.20}$$

$$\forall t \in C : x(t) > 0 \Rightarrow \mu(t, a) = 0 \text{ and } \mu(t, b) = 1. \tag{2.21}$$

The proof is in the Appendix.

2.3.3 Information Acquisition Region

The *critical types* t with $y(t) = \Pr(\alpha|\text{piv}; \sigma', n)$ are indifferent between A and B without further information, given (2.11) - (2.14). Lemma 2 shows that, for each total intensity $k = k(t) \in [0, 2]$, only types in a certain interval around the critical types acquire information.

Lemma 2. Let σ' be a strategy with $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma', n) \in (0, 1)$. When n is large enough, for any $k \in (0, 2)$ and any $g \in \{L, C\}$ there are $\phi_g^-(k) < \Pr(\alpha|\text{piv}; \sigma', n) < \phi_g^+(k)$ for such that for any best response $\sigma = (x, \mu)$ to σ' and any type $t \in g$ with $k(t) = k$,

$$x(t) > 0 \quad \Leftrightarrow \quad y(t) \in [\phi_g^-(k), \phi_g^+(k)], \tag{2.22}$$

The proof is in the Appendix. Figure 2.2 illustrates the functions ϕ_g^- and ϕ_g^+. For intuition: one can show that $y(t) \geq \phi_g^-(k)$ if and only if

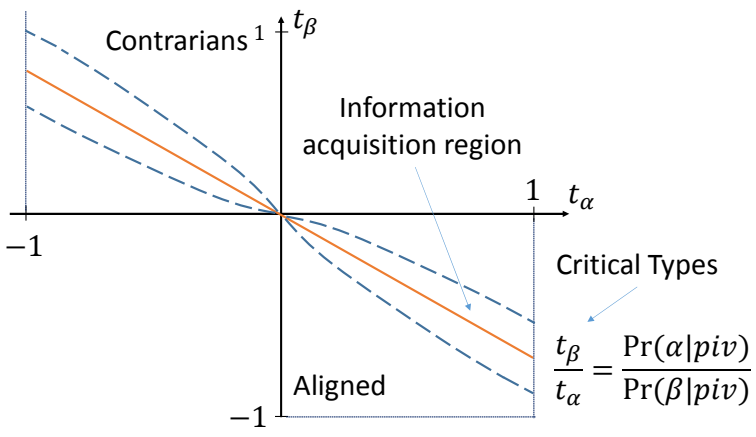
$$\frac{\Pr(\alpha|\text{piv})}{\Pr(\beta|\text{piv})} \frac{\frac{1}{2} - x^{**}(t)}{\frac{1}{2} + x^{**}(t)} \leq \frac{y(t)}{1 - y(t)}, \tag{2.23}$$

and $y(t) \leq \phi_g^+(k)$ if and only if

$$\frac{y(t)}{1 - y(t)} \leq \frac{\Pr(\alpha|\text{piv})}{\Pr(\beta|\text{piv})} \frac{\frac{1}{2} + x^{**}(t)}{\frac{1}{2} - x^{**}(t)}, \tag{2.24}$$

for $x^{**}(t; \sigma, n) = x^*(t; \sigma, n)(1 - \frac{c(x^*(y,k;\sigma,n))}{x^*(t;\sigma,n)c'(x^*(t;\sigma,n))})$, where $x^*(t; \sigma, n)$ is the solution to the first-order condition (2.18). Thus, to decide if to acquire any information, a voter discounts the precision $x^*(t; \sigma, n)$ of her optimal informative signal by a certain cost factor, and then considers if, given the discounted precision, one of the signals, a or b, sways her opinion on which policy to vote for or if none of the signals sways her.

For the later equilibrium analysis, it is key to observe that the region of types acquiring information vanishes as $n \to \infty$. This observation will allow us to understand the aggregate information acquisition of large electorates by local approximations.



Notes: The figure shows the information acquisition regions of the group g with boundaries the graphs of $\phi_g^-(k)$ and $\phi_g^+(k)$ (dashed lines).

Figure 2.2. The information acquisition regions

Lemma 3. Take any σ'. Take the best response $\sigma = (x, \mu)$. Then, for any $k \in [0, 2]$,

$$\lim_{n\to\infty} \phi_g^+(k) = \lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma', n) = \lim_{n\to\infty} \phi_g^-(k). \tag{2.25}$$

We claim that $\lim_{n\to\infty} \Pr(\text{piv}|\sigma, n) = 0$, that is, the pivotal likelihood goes to zero as $n \to \infty$. In fact, a Stirling approximation of the binomial coefficient and (2.8) yields[16] [17]

$$\Pr(\text{piv}|\omega; \sigma, n) \approx 4^n (n\pi)^{-\frac{1}{2}} \Big[q(\omega; \sigma)(1 - q(\omega; \sigma)\Big]^n, \tag{2.26}$$

and $\lim_{n\to\infty} \Pr(\text{piv}|\sigma, n) = 0$ follows from (2.26) since $q(1-q)$ is bounded above by $\frac{1}{4}$ on $[0, 1]$. Importantly, this implies

$$x^*(t; \sigma, n) \to 0, \tag{2.27}$$

given (2.19) and (2.17). Hence, $x^{**}(t; \sigma, n) \to 0$. This, together with (2.23) and (2.24) implies (2.25).

2.4 Informative Equilibrium Sequences

In the following, for the ease of the exposition, we take Φ to be strictly monotone. The results in the general case do not differ qualitatively, and are provided in Section 2.6.

16. The notation $x_n \approx y_n$ describes that two sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are *asymptotically equivalent* in the following sense: $\lim_{n\to\infty} \frac{x_n}{y_n} = 1$.

17. Stirling's formula yields $(2n)! \approx (2\pi)^{\frac{1}{2}} 2^{2n+\frac{1}{2}} n^{2n+\frac{1}{2}} e^{-2n}$ and $(n!)^2 \approx (2\pi) n^{2n+1} e^{-2n}$. Consequently, $\binom{2n}{n} \approx (2\pi)^{-\frac{1}{2}} 2^{2n+\frac{1}{2}} n^{-\frac{1}{2}} = 4^n (n\pi)^{-\frac{1}{2}}$.

We consider a sequence of elections along which the electorate's size $2n+1$ grows. For each n and a strategy σ_n, we calculate the probability that a policy $z \in \{A, B\}$ wins the support of the majority of the voters in state ω, denoted $\Pr(z|\omega; \sigma_n, n)$. We are interested in the limits of $\Pr(z|\omega; \sigma_n^*, n)$ for equilibrium sequences $(\sigma_n^*)_{n\in\mathbb{N}}$. We are particularly interested in equilibrium sequences where citizens vote in an informed manner such that the election outcomes differ across the states,

$$\lim_{n\to\infty} \Pr(A|\alpha; \sigma_n, n) \neq \lim_{n\to\infty} \Pr(A|\beta; \sigma_n, n), \tag{2.28}$$

which we call *informative*.[18]

2.4.1 Information Weighted Majority

What will matter in informative equilibria is if the aligned voters or the contrarian voters acquire more information, that is if

$$\int_{t\in L} x(t)dH(t) > \int_{t\in C} x(t)dH(t). \tag{2.29}$$

The precision $x(t)$ of a voter will play the role of an implicit *decision weight* of each voter. We will show that, in large electorates, in all states, the policy preferred by the aligned is elected when the sum of their decision weights is larger than that of the contrarians, and vice versa. A heuristic explanation is this: when all citizens acquire some information, $x(t) > 0$,

$$q(\alpha; \sigma_n^*) = \Big[\int_{t\in L} \frac{1}{2} + x(t)dH(t) + \int_{t\in C} \frac{1}{2} - x(t)dH(t)\Big], \tag{2.30}$$

$$q(\beta; \sigma_n^*) = \Big[\int_{t\in L} \frac{1}{2} - x(t)dH(t) + \int_{t\in C} \frac{1}{2} + x(t)dH(t)\Big], \tag{2.31}$$

given Lemma 1. Hence,

$$\begin{aligned} & q(\alpha; \sigma_n^*) > \frac{1}{2} > q(\beta; \sigma_n^*) \\ \Leftrightarrow \quad & \int_{t\in L} x(t)dH(t) > \int_{t\in C} x(t)dH(t). \end{aligned} \tag{2.32}$$

Thus, whenever (2.29) holds, a majority of citizens votes for A in α and B in β, that is for the outcomes prefered by the aligned. What this heuristic does not capture though is that the uninformed types, $x(t) = 0$, may play a role in the election unless they randomize their vote $50-50$.

In Section 2.4.2, we describe $\int_{t\in g} x(t)dH(t)$ in terms of the primitives of the model, thereby uncovering how the properties of a voter group g determine the

18. For large classes of settings, any typical efficiency measure, for example, full-information equivalence or utilitarian efficiency, requires equilibrium sequences to be informative.

endogenous information and the electoral power of the group. In Section 2.4.3, we state and prove the main result, characterizing all informative equilibrium sequences, thereby showing the somewhat surprising implication that the uninformed types (mis)coordinate on voting $50-50$ in the aggregate.

2.4.2 Information and Power of Voter Groups

The following result shows that, when n is large, the information $\int_{t\in g} x(t)dH(t)$ acquired by a voter group is proportional to the mass of the critical types in the voter group and proportional to a weighted mean of the intensities of these critical types. The weight of the intensities depends on the limit elasticity of the cost function, $d = \lim_{x\to 0} \frac{c'(x)x}{c(x)}$, which can be interpreted as a measure of how "cheap" information of low precision is.[19] The proof uses Lemma 3 and is provided in Section 2.4.2.1 and Section 2.4.2.2.

Lemma 4. Let $g \in \{L, C\}$. Take any strategy σ'. Let $\hat{p} = \lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma', n) \in (0,1)$, and

$$W(g,\hat{p}) = \underbrace{\Pr(\{t : t \in g\}) \Pr(\{t : y(t) = \hat{p}\}|t \in g)}_{\text{likelihood of critical types}} \underbrace{\mathrm{E}(k(t)^{\frac{2}{d-1}}|y(t) = \hat{p}, t \in g)}_{\text{weighted mean intensity of critical types}}.$$

For the best response $\sigma = (x, \mu)$ to σ',

$$\int_{t\in g} x(t)dH(t) \approx W(g,\hat{p}) \Pr(\text{piv}|\sigma', n)^{\frac{2}{d-1}} c_2, \tag{2.33}$$

where $c_2 > 0$ is a constant independent of g.

Note that, in the following, we sometimes denote types by (y,k) instead of t.

2.4.2.1 How Many Voters Acquire Information and How Much

Fix $k = k(t)$. Given Lemma 3, when n is large, only types close to critical type with $y(t) = \Pr(\alpha|\text{piv}; \sigma, n)$ acquire information, $x(t) > 0$. We show that, as a consequence, all such types choose asymptotically equivalent precisions as $n \to \infty$. In the following, we sometimes drop σ and n to shorten notation.

Claim 1. Take any strategy σ'. Take the sequence of best responses. Let $k \in [0,2]$. Take any converging sequence $(y_n)_{n\in\mathbb{N}}$. If $x(y_n,k) > 0$ for all n,

$$\frac{x_n(y_n,k)}{x_n(\Pr(\alpha|\text{piv}),k)} \approx 1. \tag{2.34}$$

19. For illustration, take e.g. $c_d(x) = x^d$. Then $\lim_{x\to 0} \frac{c_d(x)}{c_{d'}(x)} = \infty$ if $d' > d$.

Proof. Differentiating the first-order condition (2.18) implicitly, we show

$$\lim_{n\to\infty} \frac{\partial x^*(y,k;\sigma_n,n)}{\partial y} = 0. \tag{2.35}$$

in the Appendix. Together with Lemma 3, (2.35) implies (2.34). □

We show that the interval of types acquiring information, $x(t) > 0$, is asymptotically symmetric around the critical type with $y(t) = \Pr(\alpha|\text{piv})$.

Claim 2. Take any sequence σ_n'. Take the sequence of best responses $\sigma_n = (x_n, \mu_n)$. Then, for any $k \in (0,2)$,

$$\frac{x_n^{**}(\Pr(\alpha|\text{piv}),k)}{\phi_g^+(k) - \Pr(\alpha|\text{piv})} \approx \frac{x_n^{**}(\Pr(\alpha|\text{piv}),k)}{\Pr(\alpha|\text{piv}) - \phi_g^-(k)} \approx c_3, \tag{2.36}$$

for $x_n^{**}(y,k) = x_n(y,k)(1 - \frac{c(x_n(y,k))}{c'(x_n(y,k))x_n(y,k)})$, and where c_3 is a constant that only depends on $\lim_{n\to\infty} \Pr(\alpha|\text{piv})$.

Proof. The proof of Lemma 2 provides also an equivalent description of the boundary conditions (2.23) and (2.24): the information acquisition interval $[\phi_g^-(k), \phi_g^+(k)]$ is implicitly given by

$$\frac{1}{2} - x^{**}(\phi_g^-(k),k) = \chi(\phi_g^-(k)), \tag{2.37}$$

$$\frac{1}{2} + x^{**}(\phi_g^+(k),k) = \chi(\phi_g^+(k)), \tag{2.38}$$

for $\chi(y) = \frac{\Pr(\beta|\text{piv})y}{\Pr(\alpha|\text{piv})(1-y)+\Pr(\beta|\text{piv})y}$.[20] Since $\phi_g^-(k) \to \Pr(\alpha|\text{piv})$ and $\phi_g^+(k) \to \Pr(\alpha|\text{piv})$ (see Lemma 3) and since $\chi(\Pr(\alpha|\text{piv})) = \frac{1}{2}$, Taylor approximations of $\chi(\phi_g^-(k))$ and $\chi(\phi_g^-(k))$ give

$$\chi'(\Pr(\alpha|\text{piv}))\Big[\phi_g^+(k) - \Pr(\alpha|\text{piv})\Big] \approx x^{**}(\phi_g^+(k)), \tag{2.39}$$

$$\chi'(\Pr(\alpha|\text{piv}))\Big[\Pr(\alpha|\text{piv}) - \phi_g^-(k)\Big] \approx x^{**}(\phi_g^-(k)). \tag{2.40}$$

Finally, (2.36) follows from (2.39), (2.40), (2.34), and the continuity of c. □

2.4.2.2 Aggregate Information of a Voter Group

Denote by f the density of the cumulative distribution function of the threshold of doubt $y(t)$. Now, we finish the proof of Lemma 4. For this, we show that, fixing the

20. To see how (2.37) and (2.23) relate, rewrite (2.23), $\frac{\frac{1}{2}-x^{**}(t)}{\frac{1}{2}+x^{**}(t)} \leq \frac{\Pr(\beta|\text{piv})}{\Pr(\alpha|\text{piv})}\frac{y(t)}{1-y(t)}$, and rewrite further, $\frac{1}{2} - x^{**}(t) \leq \chi(y)$. Similarly, for (2.38) and (2.24).

total intensity $k = k(t)$, the average precision of citizen types is proportional to the likelihood of the critical type and the weighted intensity $k(t)^{\frac{2}{d-1}}$,

$$\begin{aligned} & \mathrm{E}(x(t)|k(t)=k, t\in g) \\ \approx\ & \underbrace{f(\Pr(\alpha|\mathrm{piv})|k(t)=k, t\in g)}_{\text{likelihood of critical type}} \underbrace{k^{\frac{2}{d-1}}}_{\textit{weighted intensity}} \Pr(\mathrm{piv})^{\frac{2}{d-1}} c_1. \end{aligned} \tag{2.41}$$

for a constant $c_1 > 0$ that only depends on $\Pr(\alpha|\mathrm{piv})$. Then, we aggregate over k to obtain (2.36). Details for this aggregation are in the Appendix.

First, given Lemma 3, Taylor approximations of the c.d.f yield

$$\frac{f(\Pr(\alpha|\mathrm{piv}) \mid k(t)=k, t\in g)\left[\phi_g^-(k) - \phi_g^+(k)\right]}{\Pr(\{t : \phi_g^-(k) \leq y(t) \leq \phi_g^+(k)\} \mid k(t)=k, t\in g)} \approx 1. \tag{2.42}$$

Combining (2.34), (2.36), and (2.42), for any k,

$$\begin{aligned} & \mathrm{E}(x(t)|k(t)=k, t\in g) \\ \approx\ & x_n(\Pr(\alpha|\mathrm{piv}), k) x_n^{**}(\Pr(\alpha|\mathrm{piv}), k) f(\Pr(\alpha|\mathrm{piv})|k(t)=k, t\in g) c_4 \\ \approx\ & x_n(\Pr(\alpha|\mathrm{piv}), k)^2 f(\Pr(\alpha|\mathrm{piv})|k(t)=k, t\in g) c_5. \end{aligned} \tag{2.43}$$

for constants $c_4 \neq 0$ and $c_5 \neq 0$ and where, for the last line, we used that $x_n^{**}(\Pr(\alpha|\mathrm{piv}), k) \approx \frac{d-1}{d} x_n(\Pr(\alpha|\mathrm{piv}), k)$ since $\frac{1}{d} = \lim_{x\to 0} \frac{c(x)}{c'(x)x}$. We see that what matters are the likelihood and the precision of the critical type. The precision of the critical type scales with the total intensity,

$$x(\Pr(\alpha|\mathrm{piv}), k)^2 \approx k(t)^{\frac{2}{d-1}} \left[\Pr(\mathrm{piv}) c_1 (\Pr(\alpha|\mathrm{piv}))\right]^{\frac{2}{d-1}}, \tag{2.44}$$

given (2.18), so that (2.44) and (2.43) imply (2.41).

2.4.2.3 Power of a Voter Group

We call

$$W(g) = W(g, \hat{p}) \tag{2.45}$$

the *power* of a voter group, where $\hat{p}$ is the unique belief $\hat{p}$ for which the electorates preferences are split, $\Phi(\hat{p}) = \frac{1}{2}$. The next lemma shows that for any informative equilibrium sequence, the threshold of doubt of the critical types converges to $\hat{p}$, so that, given Lemma 4, $W(g)$ measures the amount of information acquired by the group in any such equilibrium sequence.

Lemma 5. Let $\Phi(\hat{p}) \neq \frac{1}{2}$. Then, for any informative equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$,

$$\lim_{n\to\infty} \Pr(\alpha|\mathrm{piv}; \sigma_n^*, n) = \hat{p}. \tag{2.46}$$

The proof is provided in the Appendix. There, we show that when (2.46) does not hold, the vote shares $q(\omega;\sigma_n^*)$ do not converge to $\frac{1}{2}$, and as a consequence, the citizens choose exponentially low levels of precision. This, in turn, implies that the difference in the vote shares in α and β is exponentially small. Finally, we show that this implies that the distribution of the election outcome is asymptotically the same in both states as $n \to \infty$, which cannot be true in any informative equilibrium sequence.

2.4.3 Result

The main result shows that for all informative equilibrium sequences, the outcome preferred by the group with the larger power is elected as $n \to \infty$. Moreover, there exists an informative equilibrium sequence when information of low precision $x \approx 0$ is sufficiently cheap; this will be captured by a condition on the elasticity at zero, $\lim_{x\to 0} \frac{c'(x)x}{c(x)}$.[21] We call $W(L) \neq W(C)$, $W(L) \neq$, and $W(C) \neq 0$ the *genericity conditions*.

Theorem 1. Let $\lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$. Take any preference distribution H satisfying the genericity conditions and $\Phi(\Pr(\alpha)) \neq \frac{1}{2}$.

1. For all informative equilibrium sequences $(\sigma_n^*)_{n\in\mathbb{N}}$,

$$\begin{aligned} \lim_{n\to\infty} \Pr(A|\alpha;\sigma_n^*,n) &= \lim_{n\to\infty} \Pr(B|\beta;\sigma_n^*,n) \\ &= \begin{cases} 0 \text{ if } W(L) < W(C), \\ 1 \text{ if } W(L) > W(C). \end{cases} \end{aligned} \tag{2.47}$$

2. There is an informative equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$.

2.4.4 Proof: Power Rule

This section proves the first item of Theorem 1, showing that the order of the power of aligned and contrarians determines election outcomes. First of all, $q(\alpha;\sigma_n^*) > q(\beta;\sigma_n^*) \Leftrightarrow \frac{\int_{t\in L} x(t)dH(t)}{\int_{t\in C} x(t)dH(t)} > 1$,[22] so that, given Lemma 4, the order of $W(g)$ pins down the order of the vote shares: for n large enough,

$$q(\alpha;\sigma_n^*) > q(\beta;\sigma_n^*) \quad \Leftrightarrow \quad \frac{W(L)}{W(C)} > 1, \tag{2.48}$$

21. The same condition appears in Martinelli (2006)'s model as a sufficient condition for informative and determinate equilibrium outcomes.

22. First, partisans vote the same in both states. Second, for aligned types the likelihood to vote A in α differs by $2x(t)$ from the likelihood to vote A in β. Third, for contrarian types the likelihood to vote A in α differs by $-2x(t)$ from the likelihood to vote A in β. Together, $q(\alpha;\sigma_n^*) - q(\beta;\sigma_n^*) = \int_{t\in L} 2x(t)dH(t) - \int_{t\in C} 2x(t)dH(t)$, which implies the equivalence stated.

The key step is to establish that when the elasticity of the cost function at zero is sufficiently large, $\lim_{x\to 0} \frac{xc'(x)}{c(x)} = d > 3$, then, for any equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$ outcomes are determinate as $n \to \infty$,

$$\lim_{n\to\infty} \Pr(A|\omega; \sigma_n^*, n) \in \{0, 1\}. \tag{2.49}$$

in each state ω. For informative equilibrium sequences, this implies that A is elected in one state, and B in the other. When the vote share for policy A is higher in α than in β, A is elected in α, and B in β and vice versa when the vote share for policy A is higher in β than in α. We conclude that (2.48) and (2.49) together imply (2.47). The following section proves (2.49).

2.4.4.1 Determinate Outcomes

We show that, given $d > 3$, for any sequence of equilibria, the outcomes are determinate, as $n \to \infty$, that is, we prove (2.49).

For this, for any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ and any n, let $\boldsymbol{q}(\sigma_n) = (q(\alpha;\sigma_n), q(\beta;\sigma_n))$, and denote by $s(\omega;\boldsymbol{q}(\sigma_n)) = \Big[q(\omega;\sigma_n)(1-q(\omega;\sigma_n)(2n+1)\Big]^{\frac{1}{2}}$ the standard deviation of the vote share in ω. Let

$$\delta(\omega) = \lim_{n\to\infty} \frac{2n+1}{s(\omega;\boldsymbol{q}(\sigma_n))}\Big[q(\omega;\sigma_n) - \frac{1}{2}\Big] \tag{2.50}$$

be the normalized distance of the expected vote share to the majority threshold as $n \to \infty$.

The proof of (2.49) proceeds in three steps. The first step shows that, as a consequence of the central limit theorem, as $n \to \infty$, the asymptotic distribution of the outcome policies only depends on the distance of the vote share to the majority threshold in terms of standard deviations, i.e. $\delta(\omega)$. The proof of this step is in the Appendix.

Lemma 6. Take any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ and any state $\omega \in \{\alpha, \beta\}$. The probability that A gets elected in ω converges to

$$\lim_{n\to\infty} \Pr(A|\omega;\sigma_n) = \Phi(\delta(\omega)),$$

where $\Phi(\cdot)$ is the cumulative distribution of the standard normal distribution.

What determines the equilibrium distance of the vote shares to each other, and thereby their distance to the majority treshold, is how much information the voters acquire in equilibrium,

$$q(\alpha;\sigma_n) - q(\beta;\sigma_n) = 2\Big[\int_{t\in L} x(t)dH(t) - \int_{t\in C} x(t)dH(t)\Big]. \tag{2.51}$$

For the second step, suppose that the election is not determinate in a state ω, e.g. in α. Given Lemma 6, $\delta(\alpha) \in \mathbb{R}$. We show that $q(\alpha;\sigma_n) - q(\beta;\sigma_n)$ is of of an order larger than inverse of the standard deviation of the vote share,

$$\lim_{n\to\infty} \Big[q(\alpha;\sigma_n) - q(\beta;\sigma_n)\Big]s(\omega;\boldsymbol{q}(\sigma_n)) \in \{\infty, -\infty\}. \tag{2.52}$$

if $d > 3$. To prove (2.52), first, we show that

$$\lim_{n\to\infty} \Big[q(\alpha;\sigma_n) - q(\beta;\sigma_n)\Big] \Pr(\text{piv}|\sigma_n, n)^{-1} = \infty. \tag{2.53}$$

if $d > 3$. To see why, note that

$$\begin{aligned} &\int_{t\in L} x(t)dH(t) - \int_{t\in C} x(t)dH(t) \\ \approx\quad &\Big[W(L) - W(C)\Big] \Pr(\text{piv}|\sigma_n^*, n)^{\frac{2}{d-1}} c_2, \end{aligned} \tag{2.54}$$

given Lemma 4. Using that the pivotal likelihood goes to zero as $n \to \infty$, (2.53) follows from (2.51), (2.54), and the genericity conditions. Second, using the local central limit theorem, we show that, for all strategies with vote shares close to the majority threshold as in the lemma, the pivotal likelihood is inversely proportional to the the standard deviation of the vote share.

Lemma 7. For any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$. If $\lim_{n\to\infty} q(\omega;\sigma_n) \in (0,1)$, then

$$\lim_{n\to\infty} \Pr(\text{piv}|\omega;\sigma_n) s(\omega; \boldsymbol{q}(\sigma_n)) \quad = \quad \phi(\delta(\omega)), \tag{2.55}$$

where ϕ the probability density function of the standard normal distribution.

The proof is an application of the local central limit theorem, and provided in the Appendix.[23] Lemma 7 and (2.53) together yield (2.52).

Finally, we prove (2.49). Note that we can write $\delta(\omega) = \lim_{n\to\infty} s(\omega; \boldsymbol{q}(\sigma_n)) \Big[q(\omega;\sigma_n) - \frac{1}{2}\Big]$. Hence, (2.52) implies $\delta(\alpha) - \delta(\beta) \in \{\infty, -\infty\}$. Since $\delta(\alpha) \in \mathbb{R}$, we have $\delta(\beta) \in \{\infty, -\infty\}$. Then, Lemma 7 implies that the inference from the pivotal event is not bounded, and $\lim_{n\to\infty} \Pr(\alpha|\text{piv}, s; \sigma_n^*) = 0$ for $s \in \{a, b\}$. Hence, as $n \to \infty$, citizens vote as if they know that β holds, so that $\lim_{n\to\infty} q(\alpha;\sigma_n^*) = \Phi(0) < \frac{1}{2}$, which contradicts $\delta(\alpha) \in \mathbb{R}$. The assumption that the election outcome is determinate in β similarly leads to a contradiction.

2.4.5 Proof: Existence

This section proves existence of an informative equilibrium sequence when $\lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$. For this, first, we provide a convenient equilibrium representation.

2.4.5.1 Equilibrium Representation through Vote Shares

It follows from the analysis of the best response in Section 2.2 that, for n large enough, an equilibrium is a (non-degenerate) strategy $\sigma = (x, \mu)$ that satisfies (2.11)-(2.14), with $\sigma' = \sigma$, (2.18) for all types t with $x(t) > 0$, and (2.22).

23. See Gnedenko (1948), and Davis and McDonald (1995) for the local limit theorem for triangular arrays of integer-valued variables.

I claim that equilibrium can be alternatively characterized in terms of the vector of the expected vote shares of outcome A in state α and β, i.e.,

$$\boldsymbol{q}(\sigma) = (q(\alpha;\sigma), q(\beta;\sigma)). \tag{2.56}$$

Note that for any σ and any $\omega \in \{\alpha, \beta\}$, the vote share $q(\omega;\sigma)$ pins down the likelihood of the pivotal event conditional on ω, given (2.8). Given (2.11)-(2.14), (2.18), and (2.37)-(2.38), the vector of the pivotal likelihoods is a sufficient statistic for the best response, and therefore $\boldsymbol{q}(\sigma)$ as well. Given some vector of expected vote shares $\boldsymbol{q} = (q(\alpha), q(\beta)) \in (0,1)$, let $\sigma^{\boldsymbol{q}}$ be the best response to $\boldsymbol{q}$. Then, σ^* is an equilibrium, if and only if, $\sigma^* = \sigma^{\mathbf{q}(\sigma^*)}$. Conversely, an equilibrium can be described by a vector of vote shares $\boldsymbol{q}^*$ that is a fixed point of $\mathbf{q}(\sigma^-)$, i.e.,[24]

$$\boldsymbol{q}^* = \mathbf{q}\left(\sigma^{\boldsymbol{q}^*}\right). \tag{2.57}$$

In the following, I use the notation $\Pr(\alpha|\text{piv};\boldsymbol{q})$ to denote the posterior consistent with (2.8) and the vote shares $\boldsymbol{q}$, and also further analogous notation. The next two sections provide an analysis of the best response function $q(\sigma^-)$ in two steps. Section 2.4.5.2 describes the pivotal inference given vote shares $\boldsymbol{q}$. Section 2.4.5.3 describes the vote shares of the best response, given some pivotal inference about the state.

2.4.5.2 Inference in Large Elections

We record the intuitive fact that voters update toward the substate in which the vote share is closer to $1/2$, that is, in which the election is closer to being tied in expectation.

Lemma 8. Take any strategy σ for which $\Pr(\text{piv}|\beta;\sigma,n) \in (0,1)$. If

$$\left|q(\alpha;\sigma) - \frac{1}{2}\right| < (\leq) \left|q(\beta;\sigma) - \frac{1}{2}\right|, \tag{2.58}$$

then

$$\frac{\Pr(\text{piv}|\alpha;\sigma,n)}{\Pr(\text{piv}|\beta;\sigma,n)} > (\geq) 1. \tag{2.59}$$

Proof. The function $q(1-q)$ has an inverse u-shape on $[0,1]$ and is symmetric around its peak at $q = \frac{1}{2}$. So, $\left|q - \frac{1}{2}\right| < (\leq) \left|q' - \frac{1}{2}\right|$ implies that $q(1-q) > (\geq) q'(1-q')$. Thus, it follows from (2.8) that (2.58) implies (2.59). □

24. The ability to write an equilibrium as a finite-dimensional fixed point via (2.57) is a significant advantage. Similarly, a reduction to finite dimensional equilibrium beliefs has been useful in other settings; see Bhattacharya (2013b), Ahn and Oliveros (2012) and Heese and Lauermann (2017).

Moreover, Lemma 8 extends in an extreme form as the electorate grows large ($n \to \infty$): the event that the election is tied is infinitely more likely in the state in which the election is closer to being tied in expectation. In fact, the likelihood ratio of the pivotal event diverges exponentially fast.

Lemma 9. Consider any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$. If,

$$\lim_{n\to\infty}\left|q(\alpha;\sigma_n)-\frac{1}{2}\right| < (>)\lim_{n\to\infty}\left|q(\beta;\sigma_n)-\frac{1}{2}\right|, \tag{2.60}$$

then, for any $\kappa \geq 0$,

$$\lim_{n\to\infty}\frac{\Pr(\text{piv}|\alpha;\sigma_n,n)}{\Pr(\text{piv}|\beta;\sigma_n,n)}n^{-\kappa} = \infty(0). \tag{2.61}$$

Proof. Let

$$k_n = \frac{q(\alpha;\sigma_n)(1-q(\alpha;\sigma_n))}{q(\beta;\sigma_n)(1-q(\beta;\sigma_n))}.$$

From (2.8), the left-hand side of (2.61) is $\frac{(k_n)^n}{n^\kappa}$. The function $q(1-q)$ has an inverse u-shape on $[0,1]$ and is symmetric around its peak at $q=\frac{1}{2}$. Therefore, (2.60) implies that $\lim_{n\to\infty} k_n > 1$. So, $\lim_{n\to\infty}(k_n)^n = \infty$. Moreover, $(k_n)^n$ diverges exponentially fast and, hence, dominates the denominator n^κ, which is polynomial. □

2.4.5.3 Vote Shares and the Citizen's Inference

We show that, as $n\to\infty$, under the best response, the expected vote share for policy A in ω is given by the mass of types preferring A given the pivotal belief $\Pr(\alpha|\text{piv};\sigma'.n)$, that is $\Phi(\Pr(\alpha|\text{piv};\sigma'_n))$.

Lemma 10. Take any sequence of strategies $(\sigma'_n)_{n\in\mathbb{N}}$. Take the sequence of best responses σ_n. For any $\omega\in\{\alpha,\beta\}$,

$$\lim_{n\to\infty} q(\omega;\sigma_n) = \lim_{n\to\infty}\Phi(\Pr(\alpha|\text{piv};\sigma'_n)). \tag{2.62}$$

The proof is provided in the Appendix. The basic intuition is that, as $n\to\infty$, the precision of all types signals goes to zero uniformly, (2.27), so that, given (2.11)-(2.14), "in the limit" voters decide simply according to the pivotal belief.

2.4.5.4 Intuition: Information Acquisition can be a Complement

Lemma 10 is key to get an intuition why informative equilibrium sequences exist. The relevant economic observation coming from the lemma is that information acquisition can be complementary as a result of the pivotal inference: Take the case $\Phi(\Pr(\alpha)) > \frac{1}{2}$. Without pivotal inference, $\Pr(\alpha|\text{piv};\sigma_n,n) = \Pr(\alpha)$. Then, under the best response, A is elected with a positive margin as $n\to\infty$, $\lim_{n\to\infty} q(\omega;\sigma^*_n) > \frac{1}{2}$, using Lemma 10 and the weak law of large numbers. Since A is elected with a positive margin, the incentives to get informed are small, in fact, exponentially small, see

(2.26). However, if citizens acquire more information, so that $q(\alpha;\sigma_n)$ and $q(\beta;\sigma_n)$ differ sufficiently much, voters may make an inference about the state when conditioning on the election being tied in a way, so that $\lim_{n\to\infty}\Phi(\Pr(\alpha|\text{piv};\sigma_n,n))=\frac{1}{2}$. Then, under the best response, the election is close to being tied, thereby creating incentives to get informed. This illustrates how information acquisition can spur even more information acquisition, that is information acquisition may be complementary.

2.4.5.5 Fixed Point Argument

This section uses a fixed point argument to show that there is a sequence of equilibrium vote shares $(\boldsymbol{q}_n^*)_{n\in\mathbb{N}}$ such that the corresponding sequence of equilibrium strategies are informative. We provide the proof for the case when $\Phi(\Pr(\alpha))<\frac{1}{2}$ and when the minority group has the higher power, $W(L)<W(C)$. The proof proceeds in two steps. First, we show that for any vote share $q(\alpha)$ in α close to $\frac{1}{2}$, we find a vote share $q_n^*(\beta)$ such that the best response to $(q(\alpha),q_n^*(\beta))$ has again the same vote share in α.

Step 1. Let $\Phi(\Pr(\alpha))<\frac{1}{2}$ and $W(L)<W(C)$. For any $\epsilon>0$ small enough, any $\frac{1}{2}-\frac{\epsilon}{2}\leq q(\alpha)\leq\frac{1}{2}$, and any n large enough, there is $q_n^*(\beta)\geq\frac{1}{2}$ such that

$$q(\alpha) \quad = \quad q(\alpha;\sigma^{(q(\alpha),q_n^*(\beta))}). \tag{2.63}$$

and $q_n^*(\beta)$ is continuous in $q(\alpha)$.

Take $\frac{1}{2}-\frac{\epsilon}{2}\leq q(\alpha)\leq\frac{1}{2}$, and let $\boldsymbol{q}=(q(\alpha),q(\beta))$ in the following.

Step 1.1. If $q(\beta)=\frac{1}{2}+\epsilon$, then, for $\epsilon>0$ small enough and n large enough,

$$q(\alpha;\sigma^{\boldsymbol{q}})>q(\alpha). \tag{2.64}$$

The election is more close to being tied in α, and, by Lemma 9, voters become convinced that the state is α, i.e., $\lim_{n\to\infty}\Pr(\alpha|\text{piv};\boldsymbol{q},n)=1$. It follows from Lemma 10 that $\lim_{n\to\infty}q(\alpha;\sigma^{\boldsymbol{q}})=\Phi(1)$. Finally, (2.64) follows when ϵ is small enough since $\Phi(1)>\frac{1}{2}$.

Step 1.2. If $q(\beta)=\frac{1}{2}$, then for $\epsilon>0$ small enough and any n,

$$q(\alpha;\sigma^{\boldsymbol{q}})<q(\alpha). \tag{2.65}$$

The election is more close to being tied in β, and, by Lemma 8, voters update towards β, i.e. $\Pr(\alpha|\text{piv};\boldsymbol{q},n)\leq\Pr(\alpha)$. Since $\Phi(\Pr(\alpha))<\frac{1}{2}$, Lemma 10 implies that $\lim_{n\to\infty}q(\alpha;\sigma^{\boldsymbol{q}})<\frac{1}{2}$. Finally, (2.65) follows when ϵ is small enough.

Since $q(\alpha;\sigma^{\boldsymbol{q}})$ is continuous in $q(\beta)$, it follows from Substep 1.1, Substep 1.2, and the intermediate value theorem that, for n large enough, there is $q_n^*(\beta)$ such that (2.63) holds. It follows from the implicit function theorem that $q_n^*(\beta)$ is continuous in $q(\alpha)$.

Step 2. For any n large enough, there is $q_n^*(\alpha)$ such that

$$q_n^*(\beta) = q(\beta; \sigma^{(q_n^*(\alpha), q_n^*(\beta))}). \tag{2.66}$$

Step 2.1. For $q(\alpha) = \frac{1}{2}$, and any n large enough,

$$q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))}) > q_n^*(\beta), \tag{2.67}$$

Recall that Φ is strictly increasing. Lemma 10 together with (2.63) implies $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}_n, n) = \hat{p} \in (0,1)$ for $\boldsymbol{q}_n = (\frac{1}{2}, q_n^*(\beta))$. We claim that

$$\delta(\beta)(\boldsymbol{q}_n) \in \mathbb{R}, \tag{2.68}$$

where the notation highlights that $\delta(\beta) = \lim_{n\to\infty}(q_n^*(\beta) - \frac{1}{2})\frac{2n+1}{s(\beta;\boldsymbol{q}_n)}$ depends on $\boldsymbol{q}_n$. Otherwise, since $\delta(\alpha)(\boldsymbol{q}_n) = \lim_{n\to\infty}(q(\alpha) - \frac{1}{2})\frac{2n+1}{s(\alpha;\boldsymbol{q}_n)} = 0$, Lemma 7 implies $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}_n, n) = 1$, which contradicts the earlier observation $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}_n, n) \in (0,1)$. We claim that

$$\lim_{n\to\infty} \Big[q(\beta; \sigma^{\boldsymbol{q}_n}) - q(\alpha; \sigma^{\boldsymbol{q}_n})\Big] s(\beta; \sigma^{\boldsymbol{q}_n}) \in \{\infty, -\infty\}. \tag{2.69}$$

For this, we show that

$$\lim_{n\to\infty} \Big[q(\beta; \sigma^{\boldsymbol{q}_n}) - q(\alpha; \sigma^{\boldsymbol{q}_n})\Big] \Pr(\text{piv}|\boldsymbol{q}_n, n)^{-1} = \infty. \tag{2.70}$$

To see why, note that

$$\begin{aligned} q(\beta; \sigma^{\boldsymbol{q}_n}) - q(\alpha; \sigma^{\boldsymbol{q}_n}) &= 2\Big[\int_{t\in L} x(t)dH(t) - \int_{t\in C} x(t)dH(t)\Big] \\ &\approx \Big[W(L) - W(C)\Big] \Pr(\text{piv}|\boldsymbol{q}_n, n)^{\frac{2}{d-1}} c_2, \end{aligned} \tag{2.71}$$

where the first equality restates (2.51), and the second line follows from Lemma 4. Using that the pivotal likelihood goes to zero as $n \to \infty$, (2.70) follows from $d > 3$, (2.71), and the genericity conditions. Then, (2.69) follows from (2.70), Lemma 7 and $\delta(\omega)(\boldsymbol{q}_n) \in \mathbb{R}$. Note that $q(\alpha; \sigma^{\boldsymbol{q}_n} = \frac{1}{2}$, given (2.63), and that $q(\beta; \sigma^{\boldsymbol{q}_n}) > q(\alpha; \sigma^{\boldsymbol{q}_n})$ for n large, given (2.48) and $W(L) < W(C)$. Therefore, (2.68) and (2.69) together imply (2.67).

Step 2.2. For $q(\alpha) = \frac{1}{2} - \epsilon$, and any n large enough,

$$q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))}) < q_n^*(\beta), \tag{2.72}$$

Lemma 10 together with (2.63) implies $\lim_{n\to\infty} q(\beta; \sigma^{(q(\alpha), q^*(\beta))}) = \frac{1}{2} - \epsilon$. Since $q_n^*(\beta) > \frac{1}{2}$ by construction, (2.72) holds for n large enough.

Finally, using (2.72) and (2.67) and that $q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))})$ is continuous in $q(\alpha)$, the claim of Step 2 follows from an application of the intermediate value theorem .

It follows from Step 1 and Step 2 that for any n large enough, there is a pair of vote shares $q_n^*(\alpha)$ such that $\boldsymbol{q}_n^* = (q_n^*(\alpha), q_n^*(\beta))$ is a fixed point of $\boldsymbol{q}(\sigma^-)$. Moreover $q_n^*(\alpha) \leq \frac{1}{2} \leq q_n^*(\beta)$ by construction, implying that $\lim_{n\to\infty} \Pr(A|\alpha; \sigma^{\boldsymbol{q}}, n) \leq \frac{1}{2} \leq \lim_{n\to\infty} \Pr(A|\beta; \sigma^{\boldsymbol{q}}, n)$. Recalling that limit equilibrium outcomes are determinate when $d > 3$ (see (2.49)), this implies that the equilibrium sequence is informative. This concludes the proof when $W(C) > W(L)$ and $\Phi(\Pr(\alpha)) < \frac{1}{2}$. The other cases are analogous.

2.4.6 Weighted Welfare Rules

This section shows that for a large class of settings, elections lead to outcomes that maximize a weighted welfare rule. Roughly speaking, the result holds under independence conditions which imply that the utilities of the critical types are representative of the whole population.

Independence Conditions. We consider preference distributions for which the conditional distribution of the threshold of doubt, $F(\cdot|t \in g)$, is independent of the voter group, i.e. for all $g \in \{L, C\}$,

$$F(\cdot|t \in g) = F. \tag{2.73}$$

The conditional distribution $J(\cdot|t \in g)$ of the total intensities of types $t \in g$ is independent from F, that is, for all $g \in \{C, L\}$ and all $y \in [0, 1]$

$$J(\cdot|t \in g, y(t) = y) = J(\cdot|t \in g). \tag{2.74}$$

Recall that partisans stay uninformed and simply vote for their preferred policy, so that the information cost cannot screen their intensities. Therefore, we consider settings without partisans,[25]

$$\Pr(\{t : t \in L\}) \cup \Pr(\{t : t \in C\}) = 1. \tag{2.75}$$

Weighted Welfare. For any $\kappa \in [0, 1]$, any state ω, the κ-weighted welfare of A is

$$\sum_{i=1,\ldots,2n+1} (t_\omega(i))^\kappa, \tag{2.76}$$

where we added the label i of each citizen to the notation. The κ-weighted welfare of B is zero. Given the independence assumptions (2.73) and (2.74),

25. In fact, all results hold under the weaker condition that the welfare at stake is, in expectation, the same for A-partisans and B-partisans,

$$\begin{aligned} & \Pr(\{t : t_\alpha > 0, t_\beta > 0\}) \mathrm{E}_H(|t_\omega||\omega, \{t : t_\alpha > 0, t_\beta > 0\}) \\ = \quad & \Pr(\{t : t_\alpha < 0, t_\beta <\}) \mathrm{E}_H(|t_\omega||\omega, \{t : t_\alpha < 0, t_\beta < 0\}). \end{aligned}$$

for all $\omega \in \{\alpha, \beta\}$.

$W(L) > W(C) \Leftrightarrow \Pr(\{t : t \in L\})\mathrm{E}(k(t)^{\frac{2}{d-1}}|t \in L) > \Pr(\{t : t \in C\})\mathrm{E}(k(t)^{\frac{2}{d-1}}|t \in C)$. Using $t_\alpha = k(t)(1-y(t))$ and $t_\beta = k(t)y(t)$ and the assumption that the total intensity $k(t)$ is independent of the threshold of doubt $y(t)$,

$$\begin{aligned} W(L) &> W(C) \\ \Leftrightarrow \Pr(\{t : t \in L\})\mathrm{E}(t_\omega^{\frac{2}{d-1}}|t \in L) &> \Pr(\{t : t \in C\})\mathrm{E}(t_\omega^{\frac{2}{d-1}}|t \in C). \end{aligned} \quad (2.77)$$

for any state ω. Therefore, Theorem 1 together with the weak law of large numbers yields:

Theorem 2. Let $\lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$. Take any preference distribution H satisfying the genericity conditions and the independence conditions (2.73)- (2.75). For all informative equilibrium sequences, the elected policy maximizes κ-weighted welfare, for $\kappa = \frac{2}{d-1}$ with probability converging to 1, as $n \to \infty$.

2.5 Non-Informative Equilibrium Sequences

This section shows that there are two types of non-informative equilibrium sequences, thereby finishing the complete characterization of equilibrium sequences.

2.5.1 Voting According to the Prior is a Limit Equilibrium

There is an equilibrium sequence where, as $n \to \infty$, all citizens vote according to the prior belief. Hence, A is elected when a majority prefers A given the prior belief, $\Phi(\Pr(\alpha)) > \frac{1}{2}$, and B is elected when a majority prefers B given the prior belief, $\Phi(\Pr(\alpha)) < \frac{1}{2}$. The proof is in the Appendix.

Theorem 3. Let $\Phi(\Pr(\alpha)) \neq \frac{1}{2}$. There exists an equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$ for which

$$\lim_{n\to\infty} \Pr(A|\alpha;\sigma_n^*,n) = \lim_{n\to\infty} \Pr(A|\beta;\sigma_n^*,n) = \begin{cases} 1 \text{ if } \Phi(\Pr(\alpha)) > \frac{1}{2}, \\ 0 \text{ if } \Phi(\Pr(\alpha)) < \frac{1}{2}, \end{cases} \quad (2.78)$$

Theorem 3 and Theorem 1 show that citizens may coordinate on acquiring much information, but they may also (mis)coordinate on acquiring very few information. The proof of Theorem 3 highlights the role of the complementarity of information acquisition. Given the equilibrium sequence that converges to "voting according to the prior", citizens acquire very few information, so that the vote shares are approximately the same in each state. As a consequence, the pivotal event contains no information, $\lim_{n\to\infty} \Pr(\alpha|\text{piv};\sigma_n^*,n) = \Pr(\alpha)$, and $\lim_{n\to\infty} q(\omega;\sigma_n^*) = \Phi(\Pr(\alpha)) \neq \frac{1}{2}$ (see Lemma 10), so that either policy A or policy B wins by a clear margin. Anticipating this, citizens have in fact low incentives to get informed since the individual likelihood of affecting the outcome is exponentially small.

2.5.2 All Other Equilibria

We complete the characterization of equilibrium sequences. We show that when $\lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$, there is a third type of equilibrium sequence. This equilibrium sequence leads to the outcome that is preferred by the voter group with the larger power given the prior belief. The proof is in the Appendix.

Theorem 4. Take any preference distribution H satisfying the genericity conditions.

1. If $\lim_{x\to 0} \frac{c'(x)x}{c(x)} < 3$, all equilibrium sequences satisfy (2.78).
2. If $\lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$, there are three types of equilibrium sequences. There is an informative equilibrium sequence satisfying (2.47). There is an equilibrium satisfying (2.78), and there is an equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$ with
$$\lim_{n\to\infty} \Pr(z(\omega)|\alpha;\sigma_n^*,n) = \lim_{n\to\infty} \Pr(A|\beta;\sigma_n^*,n) = 1 \qquad (2.79)$$
where $z(\omega)$ is the outcome preferred by the group g' with the larger power, $g' = \arg\max_{g\in\{L,C\}} W(g)$.
3. Any equilibrium sequence satisfies either (2.47), (2.78), or (2.79).

The basic intuition for why there is another equilibrium sequence comes again from the observation that information acquisition of citizens can be complementary, as discussed in Section 2.4.5.3.

For illustration, let $\hat{p} > \Pr(\alpha)$, so a majority prefers B given the prior. We argue that, when the contrarians have a larger power, $W(L) < W(C)$, then, there is an equilibrium sequences where A is elected in both states. To construct such an equilibrium sequence, we employ a fixed point argument similar to the one for the informative equilibrium sequence in Section 2.4.5.5, . We show that there are equilibrium vote shares $\boldsymbol{q}_n^* = (q(\alpha)_n, q(\beta)_n)$, satisfying
$$\Phi(\Pr(\alpha)) < \frac{1}{2} < q_n(\alpha) < q_n(\beta) \qquad (2.80)$$
for n large such that $\Phi(\Pr(\alpha|\text{piv};\boldsymbol{q}_n^*)) \to \frac{1}{2}$ as $n\to\infty$. Policy A is elected in both states cince equilibrium outcomes are determinate, as $n\to\infty$, when $\lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$ (see (2.49)). Information acquisition is complementary in the following sense: citizens acquire information such that the inference about the state implies $\Phi(\Pr(\alpha|\text{piv};\boldsymbol{q}_n,n)) \approx \frac{1}{2}$. Thus, the resulting vote shares are close to $\frac{1}{2}$ by Lemma 10, making the election close to being tied, and thereby creating incentives for all citizens to acquire information.

2.6 Discussion and Extensions

2.6.1 Heterogenous Information Access and Skills

Access to information sources and the ability to interpret information vary widely across citizens. We can capture this in the alternative model where the attention

cost of the citizens depends on a private type $\gamma \in [\frac{1}{M}, M]$ for $M > 0$, and γ is drawn i.i.d. across voters from some absolutely continuous distribution with strictly positive density. For a given cost function c, a voter of *effort type* γ pays $c(\gamma, x) = \gamma c(x)$ for a signal of precision x.

It turns out that the previous analysis already captures this alternative model since cost and preference intensities are strategically equivalent: precisely, the best response of an aligned or contrarian voter with effort type γ, total intensity k and threshold of doubt y is the same as that of the voter with effort type $\gamma' = 1$, total intensity $\frac{k}{\gamma}$ and threshold of doubt y, given the characterization of the best response, (2.11)-(2.14), (2.18), (2.37) and (2.38). Therefore, it is without loss to treat the additional heterogeneity in terms of cost as part of the preference type distribution; for any distribution of γ and H, call $\hat{H}$ the *induced* preference distribution, capturing both types of heterogeneity.

When the effort type is independent of the preference types and signals of the voters, the previous welfare results (e.g. Theorem 2) carry over. This is for two reasons: first, independence implies that the policies maximizing κ-weighted welfare are the same under H and $\hat{H}$ as $n \to \infty$.[26] Second, if H satisfies the independence conditions (2.73)-(2.75), then so does $\hat{H}$.

More interesting are the situations where attention cost and preference types are correlated. It can happen that such correlation hinders welfare-efficient outcomes. An example: suppose that elder people prefer policies aligned with the state and younger people do not. Empirically, elder people care a lot about healthcare issues. Thus, suppose that it is utilitarian to choose their preferred policy. Typically, elder people are also less educated in information technologies. One can show, that, when, ceteris paribus, effort cost are much higher for the elder, their electoral power $W(L)$ is relatively low, and their preferred policy is not elected in any informative equilibrium, given $\lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$.

2.6.2 Third-Party Manipulation: Obfuscation of Voters

From the entertainment of the arena in ancient rome to hollow media campaigns on social media platforms nowadays, diverting the attention of the people from important economic and political issues, is an ubiquitous tool of politicians for managing democracies. We ask: how manipulable are elections by hollow information provision of third-parties? To analyze this question, we consider the alternative model where a third party can send a signal to specific voters, and the signal is uninformative for the issue relevant to the election ("obfuscation"). The game is as before, except that voters of the targeted group draw an uninformative signal with a given probability $\tilde{q}$, and else the costly signal with the precision $x(t)$ as acquired. Obfusca-

26. Since γ and t are independent, $\mathrm{E}((\frac{1}{\gamma}k(t)y(t))^\kappa) = \mathrm{E}(\frac{1}{\gamma}^\kappa)\mathrm{E}(k(t)y(t))^\kappa)$. Thus $\mathrm{E}((\frac{1}{\gamma}k(t)y(t))^\kappa) > 0$ is equivalent to $\mathrm{E}(k(t)y(t))^\kappa) > 0$.

tion has two effects. First, there is a *direct effect* on the precision of targeted voters; the average precision of a targeted voter choosing $x(t)$ is

$$(1-\tilde{q})x(t). \tag{2.81}$$

There is also an *indirect effect* since the targeted voter anticipates drawing an uninformative signal. This reduces the excepted benefit as well as the expected marginal benefit of her private information. One can show, that, as a consequence, a voter of a given type t is less likely to acquire any information when targeted relative to when not, and also chooses a lower precision.

Now—similar to the analysis before— the decision weight of each individual voter is given by her average precision. The *obfuscated power* of a voter group g is

$$\tilde{W}(g,\tilde{q}) = (1-\tilde{q})W(g). \tag{2.82}$$

The analogue of Theorem 1 holds: when $\lim_{x\to 0}\frac{c'(x)x}{c(x)} > 3$, in any informative equilibrium sequence, the policy preferred by the voter group with the larger power $\tilde{W}(g,\tilde{q})$ is elected. This illustrates the effectiveness of the obfuscation of voters, and implies:

Theorem 5. Let $\lim_{x\to 0}\frac{c'(x)x}{c(x)} > 3$. Take any preference distribution H satisfying the genericity conditions and $\Phi(\Pr(\alpha)) \neq \frac{1}{2}$. There is $\bar{q} < 1$, so that, if the third-party obfuscates a group g with a likelihood $\tilde{q} > \bar{q}$, then, for all informative equilibrium sequences $(\sigma_n^*)_{n\in\mathbb{N}}$,

$$\lim_{n\to\infty} \Pr(z(\omega)|\omega;\sigma_n^*,n) \quad = \quad 0$$

for all $\omega \in \{\alpha,\beta\}$, where $z(\omega)$ is the policy preferred by the obfuscated voter group in ω.

2.6.3 Polarized Preferences

This section shows that groups of voters that share common interests are less likely to win an election when the preference intensities vary more strongly across the voters in the group.

First, Lemma 11 shows that the relative power of a voter group is smaller when the preference intensities are more dispersed within the group. A preference distribution H' is a *g-intensity spread of H* if, ceteris paribus,

$$J(-|t\in g;H) \quad <_{\text{mps}} \quad J(-|t\in g;H'), \tag{2.83}$$

where $J(-|t\in g;H)$ is the conditional distribution of the (total) intensities $k(t)$ of the types $t\in g$, and where (2.83) means that $J(-|t\in g;H')$ is a mean-preserving spread of $J(-|t\in g;H)$, and by ceteris paribus, we mean that the conditional distribution of the preference types $t\in g'\neq g$ is unchanged as well as the conditional distribution of the threshold of doubt $y(t)$ of the types $t\in g$ and also the likelihood of a type being aligned or contrarian.

Lemma 11. Let $d = \lim_{x\to 0} \frac{c'(x)x}{c(x)} > 3$. Let $g \in \{C, L\}$. Take any preference distributions H, H' satisfying (2.73) - (2.75) and the genericity conditions.

1. If H' is an L-intensity spread of H,

$$\frac{W_{H'}(L)}{W_{H'}(C)} < \frac{W_H(L)}{W_H(C)}. \tag{2.84}$$

2. If H' is a C-intensity spread of H,

$$\frac{W_{H'}(L)}{W_{H'}(C)} > \frac{W_H(L)}{W_H(C)}. \tag{2.85}$$

The proof is in the Appendix. The basic argurment is that, when $d > 3$, the power of the group, $W(g)$, is proportional to the mean of a concave function of the intensities, $\mathrm{E}(k(t)^{\frac{2}{d-1}})$ see the definition in (4). The result will follow from an application of Jensen's inequality.

We lift the restriction that $t \in [-1, 1]^2$, and allow for more extreme preference types $t \in [-M, M]^2$ for $M > 0$. When M is arbitrarily large, there can be arbitrarily large within-group preference dispersion. Theorem 1 still holds, and based on it, we show that, when, ceteris paribus, the intensities within a given voter group are sufficiently dispersed, for *all* informative equilibrium sequences, the outcome preferred by the voter group is elected with probability going to 0 as $n \to \infty$. The formal statement and the proof are in the Appendix in Section 2.G.

2.6.4 Further Remarks

Median-Voter Outcomes. Whenever the contrarians have a larger power, $W(L) < W(C)$, then, the vote shares are ordered as $q(\alpha; \sigma_n^*) < q(\beta; \sigma_n^*)$ in any equilibrium when n is large, see (2.48). This implies, in particular, that the median voter-preferred policy is less likely to be elected in one of the states since the median voter prefers A only in α.

Median-Voter Theorem with Common Interests. Suppose that all voters share a common interest, $\Pr_H(\{t \in C\}) = 0$. For such situations, Theorem 1 implies that whenever information of low precision $x \approx 0$ is sufficiently cheap, $d > 3$, there is an equilibrium of the large election where the median-voter preferred outcome is elected state-by-state. In particular, outcomes are equivalent to the outcome with publicly known states ("full-information equivalence"). This has only been known for certain symmetric settings so far (Martinelli (2006), Oliveros (2013b)).

Aggregate Cost. We show that the sum of the voters' cost converges to zero in all equilibrium sequences when $\lim_{x\to 0} \frac{c'(x)x}{c(x)} \neq 3$. The proof is in the Appendix.

Lemma 12. Let $\lim_{x\to 0} \frac{c'(x)x}{c(x)} \neq 3$. Take any preference distribution satisfying the genericity conditions. Take any equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$ and let x_i be the realisation of the precision of voter $i \in \{1,\ldots,2n+1\}$. Then,

$$\lim_{n\to\infty} \Big[\sum_{i=1,\ldots,2n+1} c(x_i) \Big] = 0. \tag{2.86}$$

The lemma qualifies the discussion of welfare implications in Section 2.4.6 that does not take into account the costs of the voters.

Non-Monotone Preferences. So far, we provided the analysis assuming that the aggregate preference function Φ is strictly monotone. When Φ is non-motone, there may be multiple beliefs $\hat{p}$ for which $\Phi(\hat{p}) = \frac{1}{2}$. One can show that, for any such $\hat{p}$, there will be two equilibrium sequences, both satisfying $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma_n^*, n) = \hat{p}$. There is one informative equilibrium sequence, for which the outcome preferred by the group with the larger power $W(g,\hat{p})$ is elected state-by-state. And, there is one non-informative equilibrium sequence, for which the outcome preferred by the group with the larger power $W(g,\hat{p})$ given the prior belief, is elected; compare to Theorem 4. In particular, it may happen that different outcomes arise in different informative equilibria since the power of a voter gorup is a local notion when Φ is non-monotone.[27]

2.7 Literature

Information Aggregation Literature. This paper contributes to the literature on information aggregation in large elections. Condorcet's Jury Theorem (1785) states that if voters have common interests, but the information is dispersed throughout the electorate, then majority rule results in socially optimal outcomes. Information aggregates in the sense that electoral outcomes correspond to the choices of a fully informed welfare-maximizing social planner. Austen-Smith and Banks (1996), Feddersen and Pesendorfer (1998) have established a "modern" version of Condorcet's Jury Theorem in a setting where citizens vote strategically. Their results show that election outcomes are " full-information equivalent", that is, as if citizens have no uncertainty about the state. However, full-information equivalent outcomes are not necessarily socially optimal when voters have conflicting interests: take a situation where 51% of citizens marginally benefit from a reform, while the other 49% are severely impacted by it. This paper points at an empirical observation that has been mostly overlooked in this context: namely, that the dispersion of the voters' information is endogenous. We show how, for a large class of settings, the information

27. These results mirror known results for the model with exogenous information: if citizens were to receive a binary, conditionally i.i.d. signal about the state and Φ is non-monotone, it is known that there is a multiplicity of equilibrium sequences, some of which do not aggregate information (Bhattacharya (2013a)).

being endogenous leads to equilibria with outcomes that maximize a weighted welfare rule (Theorem 2).

This paper also contributes to the literature on elections with costly information acquisition by studying a general setup that allows the voters to have conflicting interests. Thereby, we capture many relevant economic applications; for example, distributive reforms. The previous literature has studied information aggregation in situations where all voters share a common interest.[28] For the common interest case, we generalize the result of the literature showing that information aggregation is possible under a condition on the cost function provided in Martinelli (2006). We show that the possibility result extends to general continuous preference distributions, see the discussion in Section 2.6.4. Also, we characterize all the equilibria of the voting game, revealing an equilibrium multiplicity, and establishing that, generically, information aggregation only occurs in one of three equilibria (Theorem 4).

Vote Buying and Costly Voting Literature. This paper is related to work on elections with voting cost and vote-buying. Krishna and Morgan (2011) and Krishna and Morgan (2015) have shown that elections yield first-best outcomes when voting is voluntary and costly. In a companion paper Heese (2020), we show that analogous results hold when voters have the binary choice between a costless uninformative signal and a given costly informative signal, similar to the binary choice between voting at a cost and not voting.

The model in this paper is more closely related to the literature on vote-buying. Lalley and Weyl (2018) have shown that equilibrium outcomes in large electorates are utilitarian when each voter can buy any number of votes at a total price that is quadratic in the number of votes bought. Similarly, this paper shows that when information is costly and cost are arbitrarily close to "cubic", e.g. $c(x) = x^{3+\epsilon}$, there are equilibrium sequences where limit outcomes maximize utilitarian welfare for a large class of preference distributions.[29]

Eguia and Xefteris (2018) show that vote-buying mechanisms with general price functions implement a set of weighted welfare rules. Similarly, we have shown that a subset of the same weighted welfare rules arises when political infomation is costly (Theorem 2).

Distributive Politics Literature. A rich literature in distributive politics seeks to understand if, and when, elections select policies that maximize social welfare.

28. See Martinelli (2006) and Oliveros (2013a)), and the more distantly related papers Triossi (2013) and Martinelli (2007) who study heterogeneous cost in common interest setups, and Oliveros (2013b) who studies the relationship of abstention and information cost.

29. Given the assumptions of Theorem 2, the informative equilibrium sequences lead to outcomes maximizing $\frac{1}{1+\epsilon}$-weighted welfare. Note that utilitarian welfare is what we call 1-weighted welfare. Hence, these equilibrium sequences are utilitarian except for the small set of preference distributions where the policy maximizing 1-and $\frac{1}{1+\epsilon}$-weighted welfare is not the same.

See e.g. Fernandez and Rodrik (1991), Alesina and Rodrik (1994) and Persson and Tabellini (1994). This paper introduces a novel aspect into this discussion; namely, endogenous attention to politics.[30] Fernandez and Rodrik (1991) study the effect of asymmetric information on distributive politics: there is a group of citizens who gain from a reform with certainty; however, for a majority, the individual consequences are uncertain, and given the prior, each majority voter's expected gain is negative. Without further information, this leads to rejection of the reform in a simple majority vote, even when the reform enhances the utilitarian welfare of the electorate as a whole. We would like to point out that these results may not carry over when citizens can acquire information about the distributive consequences. Future work may investigate the closer connection to this literature.

2.8 Conclusion

A modified version of the classical setting by Feddersen and Pesendorfer (1997) captures applications like distributive reforms, e.g. health care or education reforms. Election results are driven by how much demographic groups pay attention to politics. In all limit equilibria with state-dependent outcomes, the implicit decision weight of a voter is proportional to how much attention she pays to politics. This is a structural insight with wide-reaching consequences. Since citizens with higher utilities pay more attention, elections screen the voter's utilities, and the result implies strong welfare properties of elections for a large class of settings. Elections lead to policies maximizing a certain weighted welfare rule.

The results, albeit implying a positive welfare theorem when information cost are symmetric across voters, point at the scope of manipulability of elections through informational campaigns. Politicians and third parties may successfully affect elections by diverting attention of targeted groups, thereby reducing their effective electoral power. They may successfully affect elections by hampering the physical access to information, or by spreading confusion among target groups; in other words, by making it more costly to acquire knowledge about policies and their consequences. We believe that this paper can be a starting point for the analysis of many current topics concerning the role of information in elections.

30. Similar to this paper, Ali, Mihm, and Siga (2018) transports the informational approach to elections to the literature on distributive politics.

2.A Auxiliary Results

The auxiliary results are used in the proofs of this Appendix. Some of the auxiliary results will be restated as lemmas or observations in the main text when needed for the arguments there.

2.A.1 Pivotal Likelihood Ratio

For any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ and any n, let $\boldsymbol{q}(\sigma_n) = (q(\alpha;\sigma_n), q(\beta;\sigma_n))$, and denote by $s(\omega;\boldsymbol{q}(\sigma_n)) = \Big[q(\omega;\sigma_n)(1-q(\omega;\sigma_n)(2n+1)\Big]^{\frac{1}{2}}$ the standard deviation of the vote share in ω. Let

$$\delta_n(\omega) = \frac{2n+1}{s(\omega;\boldsymbol{q}(\sigma_n))}\Big[q(\omega;\sigma_n) - \frac{1}{2}\Big] \tag{2.87}$$

be the normalized distance of the expected vote share to the majority threshold, and $\delta(\omega) = \lim_{n\to\infty}\delta_n(\omega)$.

Lemma 13. For any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$,

$$\frac{\Pr(\text{piv}|\alpha;\sigma_n,n)}{\Pr(\text{piv}|\beta;\sigma_n,n)} = \Big[1 - \frac{1}{2n+1}x_n\Big]^n. \tag{2.88}$$

for

$$x_n = \frac{q(\alpha;\sigma_n)(1-q(\alpha;\sigma_n))}{q(\beta;\sigma_n)(1-q(\beta;\sigma_n))}\delta_n(\alpha)^2 - \delta_n(\beta)^2. \tag{2.89}$$

Proof. Recall the definitions of $\delta_n(\omega)$ and $s(\omega;\boldsymbol{q}(\sigma_n))$,

$$\begin{aligned}\delta_n(\omega) &= \frac{2n+1}{s(\omega;\boldsymbol{q}(\sigma_n))}(q(\omega;\sigma_n) - \frac{1}{2}) \\ &= (2n+1)^{\frac{1}{2}}\frac{q(\omega;\sigma_n) - \frac{1}{2}}{q(\omega;\sigma_n)(1-q(\omega;\sigma_n))}.\end{aligned} \tag{2.90}$$

The ratio of the likelihoods of the pivotal event in the two states is

$$\begin{aligned}&\frac{\Pr(\text{piv}|\alpha;\sigma_n,n)}{\Pr(\text{piv}|\beta;\sigma_n,n)} \\ =\ &\Big[\frac{q(\alpha;\sigma_n)(1-q(\alpha;\sigma_n)}{q(\beta;\sigma_n)(1-q(\beta;\sigma_n)}\Big]^n. \\ =\ &\Big[1 - \frac{(q(\alpha;\sigma_n)-\frac{1}{2})^2 - (q(\beta;\sigma_n)-\frac{1}{2})^2}{q(\beta;\sigma_n)(1-q(\beta;\sigma_n)}\Big]^n \\ =\ &\Big[1 - \frac{1}{2n+1}(\frac{q(\alpha;\sigma_n)(1-q(\alpha;\sigma_n))}{q(\beta;\sigma_n)(1-q(\beta;\sigma_n)}\delta_n(\alpha)^2 - \delta_n(\beta)^2\Big]^n.\end{aligned}$$

where we used (2.90) for the equality on the last line. Plugging in (2.89) yields (2.88). □

Lemma 14. Take any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$. If $\lim_{n\to\infty}\delta_n(\alpha)-\delta_n(\beta)=0$, then

$$\lim_{n\to\infty}\frac{\Pr(\text{piv}|\alpha;\sigma_n^*,n)}{\Pr(\text{piv}|\beta;\sigma_n^*,n)}=1, \tag{2.91}$$

Proof. Recalling Lemma 13, we rewrite (2.88),

$$\frac{\Pr(\text{piv}|\alpha;\sigma_n^*,n)}{\Pr(\text{piv}|\beta;\sigma_n^*,n)}=(\Big[1-\frac{1}{2n+1}x_n\Big]^n-e^{-\frac{1}{2}x_n})+e^{-\frac{1}{2}x_n} \tag{2.92}$$

with x_n given by (2.89). In the following, we analyse the two summands separately. Note that $\lim_{n\to\infty}\delta_n(\alpha)-\delta_n(\beta)=0$ implies $\lim_{n\to\infty}q(\alpha;\sigma_n)-q(\beta;\sigma_n)=0$, and therefore

$$\lim_{n\to\infty}x_n=0. \tag{2.93}$$

This yields

$$\lim_{n\to\infty}e^{-\frac{1}{2}x_n}=1, \tag{2.94}$$

Second, using the Lemmas 4.3 and 4.3 in Durrett (1991) [p.94], for all $n\in\mathbb{N}$,

$$|(1-\frac{x_n}{(2n+1)})^n-e^{-x_n}|\leq\frac{x_n^2}{(2n+1)^3} \tag{2.95}$$

Finally, (2.91) follows from (2.92) - (2.95). □

2.A.2 Proof of Lemma 6: Outcome Distribution

Lemma 6. Take any sequence of strategies $(\sigma_n)_{n\in\mathbb{N}}$ and any state $\omega\in\{\alpha,\beta\}$. The probability that A gets elected in ω converges to

$$\lim_{n\to\infty}\Pr(A|\omega;\sigma_n)=\Phi(\delta(\omega)),$$

where $\Phi(\cdot)$ is the cumulative distribution of the standard normal distribution.

Proof. Let $q_n=q(\omega,\sigma_n)$. By using the normal approximation[31]

$$\mathcal{B}(2n+1,q_n)\simeq\mathcal{N}((2n+1)q_n,(2n+1)q_n(1-q_n)), \tag{2.96}$$

31. For this normal approximation, we cannot rely on the standard central limit theorem, because q_n varies with n. Recall that for any undominated strategy, types t with $t_\alpha>0,t_\beta>0$ vote A and types t with $t_\alpha<0,t_\beta<0$ vote B. Hence, since the type distribution has a strictly positive density, there exists $\epsilon>0$ such that $\epsilon<q_n<1-\epsilon$ for all $n\in\mathbb{N}$. As a consequence, we can apply the Lindeberg-Feller central limit theorem (see Billingsley (2008), Theorem 27.2). To see why, one checks that a sufficient condition for the the Lindeberg condition is that $(2n+1)q_n(1-q_n)\to\infty$ as $n\to\infty$ since this implies that for n sufficiently large the indicator function in the condition takes the value zero.

we see that the probability that A wins the election in ω converges to

$$\Phi\left(\frac{\frac{1}{2}(2n+1)-(2n+1)\cdot q_n}{((2n+1)q_n(1-q_n))^{\frac{1}{2}}}\right). \tag{2.97}$$

Taking limits $n\to\infty$, gives

$$\begin{aligned}
&\lim_{n\to\infty}\Phi\left(\frac{\frac{1}{2}(2n+1)-(2n+1)\cdot q_n}{(2n+1)q_n(1-q_n))^{\frac{1}{2}}}\right)\\
=\;&\lim_{n\to\infty}\Phi\left(\frac{(2n+1)\frac{1}{2}-(2n+1)(\frac{1}{2}+(q_n-\frac{1}{2}))}{((2n+1)^{\frac{1}{2}}(q_n(1-q_n))^{\frac{1}{2}}}\right)\\
=\;&\lim_{n\to\infty}\Phi\left((q_n-\frac{1}{2})\Big[\frac{(2n+1)}{q_n(1-q_n)}\Big]^{\frac{1}{2}}\right)\\
=\;&\Phi(\delta(\omega)),
\end{aligned} \tag{2.98}$$

where the equalities on the last two lines hold both when $\delta(\omega)\in\{\infty,-\infty\}$ and when $\delta(\omega)\in\mathbb{R}$. For the equality on the last line, I used that the formula $s(\omega;\boldsymbol{q}(\sigma_n))=\Big[q(\omega;\sigma_n)(1-q(\omega;\sigma_n)(2n+1)\Big]^{\frac{1}{2}}$. □

2.A.3 A Lemma on the Optimal Precision

Lemma 15.

$$\lim_{n\to\infty}\frac{\partial x^*(y,k;\sigma_n,n))}{\partial y}=0. \tag{2.99}$$

uniformly for all (y,k).

Proof. Implicit differentiation of the first-order condition (2.18) shows

$$\frac{\partial x^*(y,k;\sigma_n',n))}{\partial y}=\frac{MB'(y)}{c''(x^*(y,k;\sigma_n',n))}. \tag{2.100}$$

Using (2.17) and (2.18), $MB'(y)=\Pr(\text{piv}|\sigma',n)\Big[\Pr(\beta|\text{piv};\sigma',n)-\Pr(\alpha|\text{piv};\sigma',n)\Big]=c'(x^*(y,k;\sigma_n',n))c_2$ for some constant $c_2\in\mathbb{R}$. Therefore, (2.27) together with $\lim_{x\to 0}\frac{c'(x)}{c''(x)}=\lim_{x\to 0}\frac{x}{d-1}=0$ imply (2.35). □

2.A.4 Proof of Lemma 10: Limit Vote Shares

Lemma 10. Take any sequence of strategies $(\sigma_n')_{n\in\mathbb{N}}$. Take the sequence of best responses σ_n. For any $\omega\in\{\alpha,\beta\}$,

$$\lim_{n\to\infty}q(\omega;\sigma_n)=\lim_{n\to\infty}\Phi(\Pr(\alpha|\text{piv};\sigma_n')). \tag{2.62}$$

Proof. Recall that $\Pr(\text{piv}|\sigma_n')\to 0$ as $n\to\infty$. Therefore, the first-order condition (2.18) implies that $x(t)\to 0$ uniformly. Hence, for any private signal realization s of a voter type, $\lim_{n\to\infty}\Pr(\alpha|\text{piv},s;\sigma_n^*,n)-\Pr(\alpha|\text{piv};\sigma_n^*,n)=0$. Thus, (2.11)-(2.14) imply $\lim_{n\to\infty}q(\omega;\sigma_n^*)=\Phi(\lim_{n\to\infty}\Pr(\alpha|\text{piv};\sigma_n^*))$. Finally, (2.62) follows since Φ is continuous. □

2.B Proof of Lemma 1

Since signal a is indicative of α and b of β, voters with a signal a believe state α to be more likely than voters with a signal b. In fact, given any $x > 0$, we show below that the posteriors are ordered as

$$\Pr(\alpha|b, \text{piv}; \sigma', n) < \Pr(\alpha|a, \text{piv}; \sigma', n). \tag{2.101}$$

We argue that $x(t) > 0$ implies

$$\Pr(\alpha|b, \text{piv}, \sigma', n) < y(t) < \Pr(\alpha|b, \text{piv}, \sigma', n). \tag{2.102}$$

Otherwise, given (2.11)-(2.14), there is a policy $z \in \{A, B\}$ that the voter weakly prefers, independent of her private signal $s \in \{a, b\}$. But then, she would be strictly better off by not paying for the information $x(t) > 0$ and simply voting the same after both signals. Finally, (2.11)-(2.14), and (2.102) together imply (2.20).

2.B.1 Proof of (2.101)

Note that the posterior likelihood ratio of the states conditional on a signal $s \in \{a, b\}$ with precision $x(t)$ and the event that the voter is pivotal is

$$\frac{\Pr(\alpha|s, \text{piv}; \sigma', n)}{\Pr(\beta|s, \text{piv}; \sigma', n)} = \frac{\Pr(\alpha)}{\Pr(\beta)} \frac{\Pr(\text{piv}|\alpha; \sigma', n)}{\Pr(\text{piv}|\beta; \sigma', n)} \frac{\Pr(s|\alpha; \sigma)}{\Pr(s|\beta; \sigma)}, \tag{2.103}$$

if $\Pr(\text{piv}|\beta; \sigma', n) > 0$, where I used the conditional independence of the types and signals of the other voters from the signal of the given voter. Then, the order of the likelihood ratios in (2.101) follows from $\Pr(a|\alpha; \sigma) = \frac{1}{2} + x$ and $\Pr(a|\beta; \sigma) = \frac{1}{2} - x$, and the analogous formula for $s = b$.

2.C Proof of Lemma 2

Step 1. There is $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$: take any strategy σ'. For any t, $x(t) > 0$ if and only if

$$\frac{1}{2} + x^{**}(t) \geq \chi(y(t)) \geq \frac{1}{2} - x^{**}(t) \tag{2.104}$$

for $\chi(y) = \frac{\Pr(\beta|\text{piv};\sigma',n)y}{\Pr(\alpha|\text{piv};\sigma',n)(1-y)+\Pr(\beta|\text{piv};\sigma',n)y}$ and $x^{**}(t; \sigma', n) = x^*(t; \sigma', n)(1 - \frac{c(x^*(t;\sigma',n))}{x^*(t;\sigma',n)c'(x^*(t;\sigma',n))})$, where $x^*(t; \sigma', n)$ is the unique solution to the first-order condition (2.18).

Proof. Take an aligned type. Recall that, if $x(t) > 0$, then, $x(t) = x^*(t; \sigma', n)$, and her expected utility from the policy elected in the pivotal event is given by (2.15) in α

and by (2.16) in β. Hence, an aligned type prefers choosing precision $x = x^*(t;\sigma',n)$ over voting A without further information if

$$\begin{aligned} & \Pr(\text{piv}|\sigma',n)\Big[\Pr(\alpha|\text{piv};\sigma',n)(\frac{1}{2}+x)t_\alpha + \Pr(\beta|\text{piv};\sigma',n)(\frac{1}{2}-x)t_\beta\Big] - c(x) \\ \geq\ & \Pr(\text{piv}|\sigma',n)\Big[\Pr(\alpha|\text{piv};x,\sigma',n)t_\alpha + \Pr(\beta|\text{piv};\sigma',n)t_\beta\Big]. \end{aligned} \tag{2.105}$$

Rearranging,

$$\begin{aligned} & \Pr(\text{piv}|\sigma',n)\Big[(\frac{1}{2}+x)\Big[\Pr(\alpha|\text{piv};\sigma',n)t_\alpha - \Pr(\beta|\text{piv};\sigma',n)t_\beta\Big] + \Pr(\beta|\text{piv};\sigma',n)t_\beta\Big] - c(x) \\ \geq\ & \Pr(\text{piv}|\sigma',n)\Big[\Pr(\alpha|\text{piv};\sigma',n)t_\alpha - \Pr(\beta|\text{piv};\sigma',n)t_\beta + 2\Pr(\beta|\text{piv};\sigma',n)t_\beta\Big] \end{aligned} \tag{2.106}$$

Plugging (2.17) and (2.18) into (2.106),

$$\begin{aligned} & (\frac{1}{2}+x)c'(x) - c(x) + \Pr(\text{piv}|\sigma',n)\Pr(\beta|\text{piv};\sigma')t_\beta \\ \geq\ & c'(x) + 2\Pr(\text{piv}|\sigma',n)\Pr(\beta|\text{piv};\sigma',n)t_\beta. \end{aligned} \tag{2.107}$$

We divide by $c'(x)$ rearrange, and use (2.18) and (2.17) again,

$$(\frac{1}{2}+x) - \frac{c(x)}{c'(x)} \geq 1 + \frac{\Pr(\beta|\text{piv};\sigma',n)t_\beta}{\Pr(\alpha|\text{piv};\sigma',n)t_\alpha + \Pr(\beta|\text{piv};\sigma',n)(-t_\beta)}. \tag{2.108}$$

Using $t_\alpha = k(t)(1-y(t))$ and $t_\beta = k(t)y(t)$,

$$(\frac{1}{2}+x) - \frac{c(x)}{c'(x)} \geq 1 + \frac{-\Pr(\beta|\text{piv};\sigma',n)y(t)}{\Pr(\alpha|\text{piv};\sigma',n)(1-y(t)) + \Pr(\beta|\text{piv};\sigma',n)y(t)}. \tag{2.109}$$

Rearranging gives the right inequality of (2.104). In the same way one shows that an aligned type prefers choosing precision $x = x^*(t;\sigma',n)$ over voting B without further information only if the left inequality of (2.104) holds. The argument for the contrarian types is analogous. □

Step 2. For any $g \in \{L,C\}$, any $k > 0$ and any $\epsilon > 0$, there is $\delta > 0$ such that the derivatives of

$$\frac{1}{2} + x^{**}(y,k) - \chi(y), \quad \text{and,} \tag{2.110}$$

$$\frac{1}{2} - x^{**}(y,k) - \chi(y) \tag{2.111}$$

are negative and bounded above by $-\delta$.

Proof. Since $\lim_{x\to 0}\frac{c'(x)x}{c(x)} = d$, Lemma 15 implies that the derivative of $x^{**}(y,k;\sigma',n)$ with respect to y converges to zero uniformly as $n\to\infty$. Not that χ is continuously differentiable in y; moreover, for any $\epsilon > 0$, there is $\delta > 0$

such that $\chi'(y) > \delta$ for any $y \in (\epsilon, 1-\epsilon)$ and any n.[32] For n large, enough, (2.110) and (2.111) follow. □

Now, we finish the proof of Lemma 2. Note that $\chi(\hat{y}_n) = \frac{1}{2}$ and $x^{**}(\hat{y}_n, k) > 0$ for $\hat{y}_n = \Pr(\alpha|\text{piv}; \sigma', n)$. Thus, $\chi(\hat{y}_n) < \frac{1}{2} + x^{**}(\hat{y}_n, k)$ and $\chi(\hat{y}_n) > \frac{1}{2} - x^{**}(\hat{y}_n, k)$. It follows from Step 1 and Step 2 and since $\lim_{n\to\infty} x^{**}(\hat{y}_n, k) = 0$, that, for any n large enough, there are $\phi_g^-(k), \phi_g^+(k)$ with $\phi_g^-(k) < \Pr(\alpha|\text{piv}; \sigma_n, n) < \phi_g^+(k)$ such that it is optimal to acquire information if and only if $y(t) \in [\phi_g^-(k), \phi_g^+(k)]$.

2.D Proof of Lemma 4: Aggregation over $k = k(t)$

Here, we finish the proof of Lemma 4. We have

$$\begin{aligned}\int_{t\in g} x(t)dH(t) &= \Pr(t \in g)\mathrm{E}(x(t)|t \in g)\\ &= \Pr(t \in g)\mathrm{E}(\mathrm{E}(x(t)|t \in g, k(t) = k))\\ &= \Pr(t \in g)\int_{k=k(t)} \mathrm{E}(x(t)|t \in g, k(t) = k)dH(k(t)|t \in g), \end{aligned} \tag{2.112}$$

where we used the law of iterated expectations for the second equality and where $H(k(t)|t \in g)$ is the conditional distribution of the total intensity of the types $t \in g$. Using (2.41),

$$\begin{aligned}&\int_{t\in g} x(t)dH(t)\\ \approx\ &\Pr(t \in g)\int_{k=k(t)} f(\Pr(\alpha|\text{piv})|k(t) = k, t \in g)k^{\frac{2}{d-1}}dH(k(t)|t \in g)\\ &\Pr(\text{piv})^{\frac{2}{d-1}}c_2 \end{aligned} \tag{2.113}$$

for a constant $c_2 \neq 0$ that only depends on $\Pr(\alpha|\text{piv})$. Rewriting,

$$\begin{aligned}&\int_{t\in g} x(t)dH(t)\\ \approx\ &\Pr(t \in g)f(\Pr(\alpha|\text{piv})|t \in g)\mathrm{E}\Big[k(t)^{\frac{2}{d-1}}|t \in g, y(t) = \Pr(\alpha|\text{piv})\Big] \end{aligned} \tag{2.114}$$

Taking limits $n \to \infty$,

$$\int_{t\in g} x(t)dH(t) \approx W(g, \hat{p})\Pr(\text{piv})^{\frac{2}{d-1}}c_2, \tag{2.115}$$

for $\hat{p} = \lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma, n)$.

32. For any $p \in (0,1)$, $\frac{\partial}{\partial y}(\frac{py}{py+(1-p)(1-y)}) = \frac{(1-p)p}{(p(2y-1)-y+1)^2}$. Thus, for any $\epsilon > 0$, there is $\delta > 0$ such that for all $p \in (\epsilon, 1-\epsilon)$, $\frac{\partial}{\partial y}(\frac{py}{py+(1-p)(1-y)}) > \delta$. The assumption $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma', n) \in (0,1)$ implies that, moreover, there is $\delta > 0$ such that $\chi'(y)$ is uniformly bounded below by a positive constant for any n large enough.

2.E Proof of Lemma 5

Suppose that $\lim_{n\to\infty} \Pr(\alpha|\text{piv};\sigma_n^*) \neq \hat{p}$. Then, Lemma 10 implies $\lim_{n\to\infty} q(\omega;\sigma_n) \neq \frac{1}{2}$ for $\omega \in \{\alpha,\beta\}$. Then, (2.26) implies that the pivotal likelihood is exponentially small, which in turn implies that $x(t)$ is exponentially small for all t, given (2.19) and (2.17). Therefore, the difference in the vote shares is exponentially small, which implies $\left[q(\alpha;\sigma_n^*) - q(\beta;\sigma_n^*\right]s(\omega;\sigma_n^*) = 0$ for $\omega \in \{\alpha,\beta\}$ since the standard deviation of the realized votes is of order $n^{\frac{1}{2}}$. Hence $\delta(\alpha) = \delta(\beta)$. Finally, Lemma 6 implies $\lim_{n\to\infty} \Pr(A|\alpha;\sigma_n^*) = \lim_{n\to\infty} \Pr(A|\beta;\sigma_n^*)$. But this contradicts with the assumption that $(\sigma_n^*)_{n\in\mathbb{N}}$ is an informative equilibrium sequence.

2.F Proof of Lemma 7

Fix a voter and a state ω. The number of realized A-votes among the votes of the other citizens is the sum of $2n$ i.i.d. Bernoulli variables with mean $q(\omega;\sigma_n)$. Let $X_{k,n} = \mathscr{B}(1, q(\omega;\sigma_n))$ for any $1 \leq k \leq 2n$ and $n \in \mathbb{N}$. Recall the assumption $\lim_{n\to\infty} q(\omega;\sigma_n) \in (0,1)$, and check that the conditions of Theorem 2 in Davis and McDonald (1995) are satisfied for $X_{k,n}$, $a_n = 2nq(\omega;\sigma_n)$, and $b_n = \Big[q(\omega;\sigma_n)(1 - q(\omega;\sigma_n)\Big]^{\frac{1}{2}})(2n)^{\frac{1}{2}}$. Note that $b_n \approx s(\omega;\boldsymbol{q}(\sigma_n))$. Further note that $\Pr(\text{piv}|\omega;\sigma_n,n) = \Pr(T_n = n)$ for $T_n = \sum_{i=1,\ldots,2n} X_{k,n}$. Application of Theorem 2 in Davis and McDonald (1995) gives

$$\lim_{n\to\infty} \Pr(\text{piv}|\omega;\sigma_n,n)s(\omega,\sigma_n) = \phi(\delta(\omega)). \tag{2.116}$$

2.G Proof of Lemma 11

Proof. Recall $\lim_{x\to 0} \frac{c'(x)x}{c(x)} = d$. Since $k^{\frac{2}{d-1}}$ is strictly concave when $d > 3$, an application of Jensen's inequality shows that for any g-intensity spread H' of H,

$$\mathrm{E}(k^{\frac{2}{d-1}}|t \in g;H') < \mathrm{E}(k^{\frac{2}{d-1}}|t \in g;H) \tag{2.117}$$

It follows from the definition of a g-intensity spread that for $g \neq g' \in \{L,C\}$,

$$\mathrm{E}(k^{\frac{2}{d-1}}|t \in g';H') = \mathrm{E}(k^{\frac{2}{d-1}}|t \in g';H). \tag{2.118}$$

Since H and H' satisfy (2.74), $\mathrm{E}(k^{\frac{2}{d-1}}|t \in g';H') = \mathrm{E}(k^{\frac{2}{d-1}}|t \in g', y(t) = \hat{p};H)$ and $\mathrm{E}(k^{\frac{2}{d-1}}|t \in g';H') = \mathrm{E}(k^{\frac{2}{d-1}}|t \in g', y(t) = \hat{p};H)$ for all $g' \in \{L,C\}$. Therefore (2.117), (2.118), the definition of $W(g)$ (see (2.45)) and the definition a g-intensity spread together imply (2.84) for $g = L$ and (2.85) for $g = C$, which finishes the proof of the lemma. □

Theorem 6. Let $\lim_{x\to 0}\frac{c'(x)x}{c(x)} > 3$. Let $g \in \{L, C\}$. Take any preference distribution H satisfying the genericity conditions and the independence conditions (2.73) - (2.75). When M is large enough, there is a g-intensity spread H' of H such that

$$\lim_{n\to\infty} \Pr(z(\omega)|\omega; \sigma_n^*, n) \quad = \quad 0 \tag{2.119}$$

for all $\omega \in \{\alpha, \beta\}$, where $z(\omega)$ is the policy preferred by the voter group g in ω.

Consider the case $g = L$. Given Lemma 1 and Lemma 4, it remains to show that for any H there is an L-intensity spread H', so that

$$\frac{W_{H'}(L)}{W_{H'}(C)} < 1. \tag{2.120}$$

For this, it suffices to show that for any ϵ, we can choose H', so that

$$\mathrm{E}_{H'}(k(t)^{\frac{2}{d-1}}|t \in g) < \epsilon \tag{2.121}$$

since the genericity conditions ensure that $W_H(C) = W_{H'}(C) > 0$. Take L-intensity spreads $H'(\kappa)$ of H, so that

$$\begin{aligned} &\Pr(\{t : \kappa \leq k(t) \leq \kappa + \delta\}|t \in L; H'(\kappa)) \\ &+\Pr(\{t : 0 \leq k(t) \leq \delta\}|t \in L; H'(\kappa)) \geq 1 - \delta \end{aligned} \tag{2.122}$$

for some $\kappa > 0$ and $\delta > 0$. Since the mean of the intensities is preserved under the L-intensity spread, the iterated law of expectation gives $\lim_{\delta\to 0}\Pr(\{t : \kappa \leq k(t) \leq \kappa + \delta\}|t \in L; H'(\kappa))\kappa = \mathrm{E}(k(t)|t \in L; H'(\kappa))$. Hence,

$$\begin{aligned} \lim_{\delta\to 0}\mathrm{E}(k(t)^{\frac{2}{d-1}}|t \in L; H'(\kappa)) \quad &= \quad \lim_{\delta\to 0}\Pr(\{t : \kappa \leq k(t) \leq \kappa + \delta\}|t \in L; H'(\kappa))\kappa^{\frac{2}{d-1}} \\ &= \quad \frac{\mathrm{E}(k(t)|t \in L; H'(\kappa))}{\kappa}\kappa^{\frac{2}{d-1}} \overset{\kappa\to\infty}{\to} 0, \end{aligned} \tag{2.123}$$

where I used that $d > 3$ and hence $\frac{2}{d-1} < 1$. We conclude that for κ large enough and $\kappa < M$, we find an L-intensity spread of H, so that (2.120) holds. This finishes the proof for $g = L$. The proof for $g = C$ is analogous.

2.H Proof of Theorem 3

Recall that equilibrium can be alternatively characterized in terms of the vector of the expected vote shares of outcome A in state α and β, (2.56). Let

$$\mathrm{Q}_{\epsilon,n} = \{\boldsymbol{q} = (q(\alpha), q(\beta)) : |\boldsymbol{q} - (\frac{1}{2}, \frac{1}{2})| > \epsilon \text{ and } |q(\alpha) - q(\beta)| < \frac{1}{n^2}\} \tag{2.124}$$

We claim that when δ is small enough and n large enough, the best response is a self-map on $\mathrm{B}_{\delta,n}$,

$$\boldsymbol{q} \in \mathrm{Q}_{\epsilon,n} \Rightarrow \boldsymbol{q}(\sigma^{\boldsymbol{q}}) \in \mathrm{Q}_{\epsilon,n}. \tag{2.125}$$

The proof consists of three steps: Take $q \in Q_{\epsilon,n}$. First, the vote shares in the two states are almost identical; in particular, the probability of a tie is also almost the same in the two states. Therefore, the pivotal event contains no information as $n \to \infty$,

$$\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma^{\boldsymbol{q}}, n) = \Pr(\alpha), \tag{2.126}$$

To see why, recall that for any $\boldsymbol{q} \in Q_{\epsilon,n}$, $q(\alpha) - q(\beta)| \leq \frac{1}{n^2}$. Recalling (2.90), this implies $\lim_{n\to\infty} \delta_n(\alpha) - \delta_n(\beta) = 0$. Then, (2.126) follows from Lemma 14. Using Lemma 10, (2.126) implies $|q(\sigma^{\boldsymbol{q}}) - (\frac{1}{2}, \frac{1}{2})| > \delta$ when ϵ is small enough and n large enough.

Second, the likelihood of the pivotal event is exponentially small, given (2.26). Thus, also $q(\alpha; \sigma^{\boldsymbol{q}}) - q(\beta; \sigma^{\boldsymbol{q}})$ is exponentially small, given Lemma 4.

Finally, an application of Kakutani's fixed point theorem shows that there is a sequence of equilibrium vote shares $(\boldsymbol{q}_n^*)_{n\in\mathbb{N}}$, that is, vote shares satisfying (2.56), and, given (2.126) and Lemma 10,

$$\lim_{n\to\infty} q_n^*(\omega) = \Phi(\Pr(\alpha)). \tag{2.127}$$

for all states ω. The theorem follows from the weak law of large numbers and $\Phi(\Pr(\alpha)) \neq \frac{1}{2}$.

2.I Proof of Theorem 4

2.I.1 Third Item of Theorem 4

Take any equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$. with $\hat{p} = \lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma_n^*, n)$. Given (2.48), the order of the vote shares is pinned down by the order of the voter groups' power $W(g)$. Inspection of the cases shows that it is sufficient to show that when $W(L) > W(C)$, there is no equilibrium sequence for which, in both states, the outcome preferred by the contrarians is elected given the prior belief.

Case 3. $\Phi(\Pr(\alpha)) > \frac{1}{2}$.

Suppose $\lim_{n\to\infty} \Pr(A|\alpha; \sigma_n^*, n) = \lim_{n\to\infty} \Pr(A|\alpha; \sigma_n^*, n) = 0$. Hence, $q(\omega; \sigma_n^*) \leq \frac{1}{2}$ for n large. The order $W(L) > W(C)$ pins down the order of the vote shares, $q(\alpha; \sigma_n^*) > q(\beta; \sigma_n^*)$ for n large. Thus, $\Pr(\text{piv}|\alpha; \sigma_n^*, n) \geq \Pr(\text{piv}|\beta; \sigma_n^*, n)$ for n large enough. Since Φ is strictly increasing, $\lim_{n\to\infty} \Phi(\Pr(\text{piv}|\alpha; \sigma_n^*, n)) > \Phi(\Pr(\alpha))$. Lemma 10 implies $\lim_{n\to\infty} q(\omega; \sigma_n^*) > \frac{1}{2}$. The weak law of large numbers implies $\lim_{n\to\infty} \Pr(A|\alpha; \sigma_n^*, n) = \lim_{n\to\infty} \Pr(A|\alpha; \sigma_n^*, n) = 1$, contradicting the initial assumption.

Case 4. $\Phi(\Pr(\alpha)) < \frac{1}{2}$.

The proof is analogous to the case $\Phi(\Pr(\alpha)) > \frac{1}{2}$.

2.I.2 First Item of Theorem 4

Take any equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$.

Case 1. $\lim_{n\to\infty} \Phi(\Pr(\alpha|\text{piv};\sigma_n^*,n)) \neq \frac{1}{2}$

Given (2.26), the likelihood of the pivotal event is exponentially small. As a consequence, the difference of the vote shares $q(\alpha;\sigma_n^*) - q(\beta;\sigma_n^*)$ is exponentially small, given Lemma 4 and (2.51). This implies $\delta_n(\alpha) - \delta_n(\beta) \to 0$ (see the definition (2.87)). It follows from Lemma 14 that $\lim_{n\to\infty} \Pr(\alpha|\text{piv};\sigma_n^*,n) = \Pr(\alpha)$. Then, it follows from the weak law of large numbers that the equilibrium sequence satisfies (2.78). This was to be shown.

Case 2. $\lim_{n\to\infty} \Phi(\Pr(\alpha|\text{piv};\sigma_n^*,n)) = \frac{1}{2}$

Recall that $\hat{p}$ is the unique belief with $\Phi(\hat{p}) = \frac{1}{2}$, thus $\hat{p} = \lim_{n\to\infty} \Pr(\alpha|\text{piv};\sigma_n^*,n)$. Recall the definition of $\delta_n(\omega)$, that is (2.90). We show that

$$\lim_{n\to\infty} \delta_n(\alpha) - \delta_n(\beta) = 0. \tag{2.128}$$

For this, first, we show

$$\lim_{n\to\infty} \Big[q(\alpha;\sigma_n) - q(\beta;\sigma_n)\Big] \Pr(\text{piv}|\sigma_n,n)^{-1} = 0. \tag{2.129}$$

if $d < 3$. To see why, note that

$$\begin{aligned} &\int_{t\in L} x(t)dH(t) - \int_{t\in C} x(t)dH(t) \\ \approx\ &\Big[W(L) - W(C)\Big] \Pr(\text{piv}|\sigma_n^*,n)^{\frac{2}{d-1}} c_2, \end{aligned} \tag{2.130}$$

given Lemma 4. Using that the pivotal likelihood goes to zero as $n \to \infty$, (2.129) follows from (2.51), (2.130) and $d < 3$. Given Lemma 7, the pivotal likelihood is of an order weakly smaller than $s(\omega;\sigma_n^*)^{-1}$. Hence, (2.129) implies $\lim_{n\to\infty} \Big[q(\alpha;\sigma_n) - q(\beta;\sigma_n)\Big] s(\omega;\sigma_n^*) = 0$, and thereby (2.128).

Now, Lemma 14 and (2.128) imply $\lim_{n\to\infty} \Pr(\alpha|\text{piv};\sigma_n^*) = \Pr(\alpha)$. However, this yields a contradiction to $\lim_{n\to\infty} \Phi(\Pr(\alpha|\text{piv};\sigma_n^*) = \frac{1}{2}$ since $\Phi(\Pr(\alpha)) \neq \frac{1}{2}$ by assumption. Hence, all equilibrium sequences satisfy the condition of Case (1), and we have already shown that this condition implies (2.78), which was to be shown.

2.I.3 Second Item of Theorem 4

This section uses a fixed point argument to show that there is a sequence of equilibrium vote shares $(\boldsymbol{q}_n^*)_{n\in\mathbb{N}}$ such that the corresponding sequence of equilibrium strategies satisfies (2.79). We provide the proof for the case when $\Phi(\Pr(\alpha)) < \frac{1}{2}$ and when the minority group has the higher power, $W(L) < W(C)$. The proof proceeds in two steps. First, we show that for any vote share $q(\alpha)$ in α close to $\frac{1}{2}$, we find a vote share $q_n^*(\beta)$ such that the best response to $\boldsymbol{q} = (q(\alpha), q_n^*(\beta))$ has again the same vote share in α.

Step 1. Let $\Phi(\Pr(\alpha)) < \frac{1}{2}$ and $W(L) < W(C)$. For any $\epsilon > 0$ small enough, any $\frac{1}{2} \leq q(\alpha) \leq \frac{1}{2} + \frac{\epsilon}{2}$, and any n large enough, there is $q_n^*(\beta) \geq \frac{1}{2}$ such that

$$q(\alpha) \quad = \quad q(\alpha; \sigma^{(q(\alpha), q_n^*(\beta))}). \tag{2.131}$$

and $q_n^*(\beta)$ is continuous in $q(\alpha)$.

Let $\boldsymbol{q} = (q(\alpha), q(\beta))$ in the following.

Step 1.1. If $q(\beta) = \frac{1}{2} + \epsilon$, then, for ϵ small enough and n large enough,

$$q(\alpha; \sigma^{\boldsymbol{q}}) > q(\alpha). \tag{2.132}$$

The election is more close to being tied in α, and, by Lemma 2, voters become convinced that the state is α, i.e., $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}, n) = 1$. It follows from Lemma 10 that $\lim_{n\to\infty} q(\alpha; \sigma^{\boldsymbol{q}}) = \Phi(1)$. Finally, (2.64) follows when ϵ is small enough since $\Phi(1) > \frac{1}{2}$.

Step 1.2. If $q(\beta) = \frac{1}{2}$, then for ϵ small enough and any n,

$$q(\alpha; \sigma^{\boldsymbol{q}}) < q(\alpha). \tag{2.133}$$

The election is more close to being tied in β, and, by Lemma 1, voters update towards β, i.e. $\Pr(\alpha|\text{piv}; \boldsymbol{q}, n) \leq \Pr(\alpha)$. Since $\Phi(\Pr(\alpha)) < \frac{1}{2}$, Lemma 10 implies that $\lim_{n\to\infty} q(\alpha; \sigma^{\boldsymbol{q}}) < \frac{1}{2}$. Finally, (2.65) follows when ϵ is small enough.

Since $q(\alpha; \sigma^{\boldsymbol{q}})$ is continuous in $q(\beta)$, it follows from Step 1.1, Step 1.2, and the intermediate value theorem that, for n large enough, there is $q_n^*(\beta)$ such that (2.131) holds. It follows from the implicit function theorem that $q_n^*(\beta)$ is continuous in $q(\alpha)$.

Step 2. For any n large enough, there is $q_n^*(\alpha)$ such that

$$q_n^*(\beta) = q(\beta; \sigma^{(q_n^*(\alpha), q_n^*(\beta))}). \tag{2.134}$$

Step 2.1. For $q(\alpha) = \frac{1}{2}$, and any n large enough,

$$q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))}) > q_n^*(\beta), \tag{2.135}$$

Recall that Φ is strictly increasing. Lemma 10 together with (2.131) implies $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}_n, n) = \hat{p} \in (0, 1)$ for $\boldsymbol{q}_n = (\frac{1}{2}, q_n^*(\beta))$. We claim that

$$\delta(\beta)(\sigma^{\boldsymbol{q}_n}) = \lim_{n\to\infty}(q_n^*(\beta) - \frac{1}{2})s(\beta; \sigma^{\boldsymbol{q}_n}) \in \mathbb{R}. \tag{2.136}$$

Otherwise, since $\delta(\alpha)(\sigma^{\boldsymbol{q}_n}) = \lim_{n\to\infty}(q(\alpha) - \frac{1}{2})s(\beta; \sigma^{\boldsymbol{q}_n}) = 0$, Lemma 7 implies $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}_n, n) = 1$, which contradicts the earlier observation $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}_n, n) \in (0, 1)$. Recall (2.53); together with Lemma 7 and $\delta(\omega)(\sigma^{\boldsymbol{q}_n}) \in \mathbb{R}$ for $\omega \in \{\alpha, \beta\}$,

$$\lim_{n\to\infty} \left[q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))}) - q(\alpha; \sigma^{(q(\alpha), q_n^*(\beta))}\right] s(\beta; \sigma^{\boldsymbol{q}_n}) \in \{\infty, -\infty\}. \tag{2.137}$$

Since $q(\alpha; \sigma^{(q(\alpha), q_n^*(\beta))}) = \frac{1}{2}$, given (2.63), and since $q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))} > q(\alpha; \sigma^{(q(\alpha), q_n^*(\beta))}$ for n large, given (2.48) and $W(L) < W(C)$, (2.136) and (2.137) together imply (2.135).

Step 2.2. For $q(\alpha) = \frac{1}{2} + \epsilon$, and any n large enough,

$$q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))}) < q_n^*(\beta), \tag{2.138}$$

Recall Lemma 10, which states $\lim_{n\to\infty} q(\omega; \sigma^{(q(\alpha), q^*(\beta))}) = \lim_{n\to\infty} \Phi(\Pr(\alpha|\text{piv}; \sigma_n, n))$. Given (2.131), $lim_{n\to\infty}\Phi(\Pr(\alpha|\text{piv}; \sigma_n, n)) = \frac{1}{2} + \epsilon$. Since Φ is strictly increasing, this implies $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \boldsymbol{q}_n) > \Pr(\alpha)$, given that $\Phi(\Pr(\alpha)) < \frac{1}{2}$. Recalling Lemma 13, this implies that

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} \frac{q(\alpha)(1-q(\alpha)}{(q(\beta)(1-q(\beta))}\delta_n(\alpha; \boldsymbol{q}_n)^2 - \delta_n(\beta; \boldsymbol{q}_n)^2 \in (0,1) \tag{2.139}$$

Note that, in particular, this implies $q_n^*(\beta) \to \frac{1}{2} + \epsilon$. Now, we study the best response $\sigma^{\boldsymbol{q}_n}$. The pivotal likelihood given $\boldsymbol{q}_n$ is exponentially small since $\lim_{n\to\infty} q_n^*(\beta) = q(\alpha) = \frac{1}{2}$ and (2.26). Hence, given Lemma 4 and (2.51),

$$q(\alpha; \sigma^{\boldsymbol{q}_n}) - q(\beta; \sigma^{\boldsymbol{q}_n}) \leq y^n \tag{2.140}$$

for some $0 < y < 1$. This together with (2.139) implies (2.138) for n large enough.

Finally, using (2.135) and (2.138) and that $q(\beta; \sigma^{(q(\alpha), q_n^*(\beta))})$ is continuous in $q(\alpha)$, the intermediate value theorem implies Step 2.

It follows from Step 1 and Step 2 that for any n large enough, there is a pair of vote shares $q_n^*(\alpha)$ such that $\boldsymbol{q}_n^* = (q_n^*(\alpha), q_n^*(\beta))$ is a fixed point of $\boldsymbol{q}(\sigma^-)$. Moreover $q_n^*(\alpha) \leq \frac{1}{2} \leq q_n^*(\beta)$ by construction, implying that $\lim_{n\to\infty} \Pr(A|\alpha; \sigma^{\boldsymbol{q}}, n) \leq \frac{1}{2} \leq \lim_{n\to\infty} \Pr(A|\beta; \sigma^{\boldsymbol{q}}, n)$. Recalling from (2.49) that limit equilibrium outcomes are determinate when $d > 3$, this implies that the equilibrium sequence is informative. This concludes the proof of existence of informative equilibrium sequences when $W(C) > W(L)$ and $\Phi(\Pr(\alpha)) < \frac{1}{2}$. The proof for the other cases is analogous.

2.J Proof of Lemma 12

Case 1. $\lim_{n\to\infty} \Phi(\Pr(\alpha|\text{piv}; \sigma_n^*)) \neq \frac{1}{2}$.

Then, (2.26) implies that the likelihood of the pivotal event is exponentially small,

$$\Pr(\text{piv}|\alpha; \sigma_n^*) < z^n \tag{2.141}$$

for some $0 < z < 1$ and for n large enough. Hence, for all t,

$$x^*(t) < z^n c_2 \tag{2.142}$$

for some $c_2 \in \mathbb{R}$, given (2.19). Finally, this implies that $(2n+1)\text{E}(c(x(t)) \to 0$ as $n \to \infty$ since $c(x)$ is approximately polynomial for x small enough, given (2.1). An application of the weak law of large number shows that the realized sum of the votes converges to 0 as $n \to \infty$.

Case 2. $\lim_{n\to\infty} \Pr(\alpha|\text{piv}; \sigma_n^*, n) = \hat{p}$.

Recall that $\Phi(\hat{p}) = \frac{1}{2}$. Fix $g \in \{\ell, s\}$. We use the notation $(y,k) = (y(t), k(t))$ for types $t \in g$, noting that (y,k) pin down the type uniquely. Let $\alpha = \arg\max(|q(\alpha;\sigma_n^*) - \frac{1}{2}|, |q(\alpha;\sigma_n^*) - \frac{1}{2}|)$. The other case will be analogous. First,

$$\lim_{n\to\infty} \frac{\Pr(\text{piv}|\alpha;\sigma_n^*)}{\Pr(\text{piv}|\beta;\sigma_n^*)} = \frac{\Pr(\beta)}{\Pr(\alpha)} \frac{\hat{p}}{1-\hat{p}}. \tag{2.143}$$

Multiplication of the first-order condition (2.18) by $n^{\frac{1}{2}}$ together with (2.143) yields

$$\begin{aligned} & n^{\frac{1}{2}} c'(x^*(\Pr(\alpha|\text{piv};\sigma_n^*, n), 1)) \\ = \;& n^{\frac{1}{2}} \Pr(\alpha) \Pr(\text{piv}|\alpha;\sigma_n^*, n) k(t) \Big[(1-y(t)) - y(t) \frac{\Pr(\text{piv}|\beta;\sigma_n^*, n)}{\Pr(\text{piv}|\alpha;\sigma_n^*, n)}\Big] \end{aligned} \tag{2.144}$$

Note that

$$\begin{aligned} 4^n (q(1-q))^n &= 4^n \Big[(\frac{1}{2} - (\frac{1}{2} - q))(\frac{1}{2} + (\frac{1}{2} - q)\Big]^n \\ &= 4^n (\frac{1}{4} - (\frac{1}{2} - q)^2)^n \\ &= (1 - 4\frac{(n^{\frac{1}{2}}(\frac{1}{2} - q))^2}{n})^n. \end{aligned} \tag{2.145}$$

for all $q \in (0,1)$. Combining (2.26) with (2.144) and (2.145) gives

$$n^{\frac{1}{2}} c'(x^*(\Pr(\alpha|\text{piv};\sigma_n^*, n), 1)) \approx c_2 (1 - 4\frac{(n^{\frac{1}{2}}(\frac{1}{2} - q(\alpha;\sigma_n^*))^2}{n})^n$$

for some constant $c_2 > 0$. Multiplication of both sides with $\delta_n = n^{\frac{1}{2}}|q(\alpha;\sigma_n^*) - \frac{1}{2}|$ yields

$$\delta_n n^{\frac{1}{2}} c'(x^*(\Pr(\alpha|\text{piv};\sigma_n^*, n), 1)) \approx c_3 \delta_n e^{-4\delta_n^2} + \delta_n \Big[(1 - 4\frac{\delta_n^2}{n})^n - e^{-4\delta_n^2}\Big]. \tag{2.146}$$

for some constant $c_3 > 0$. Using Lemmas 4.3 and 4.3 in Durrett (1991),

$$(1 - 4\frac{\delta_n^2}{n})^n - e^{-4\delta_n^2} \leq \frac{16\delta_n^4}{n^3}. \tag{2.147}$$

Therefore, $\lim_{n\to\infty} \delta_n \Big[(1 - 4\frac{\delta_n^2}{n})^n - e^{4\delta_n}\Big] = 0$. Since, given $d > 3$, all equilibrium sequences are determinate by (2.49), Lemma 6 implies $\lim_{n\to\infty} \delta_n = \infty$, which in turn implies $\lim_{n\to\infty} \delta_n e^{-4\delta_n} = 0$. I conclude,

$$\lim_{n\to\infty} \delta_n n^{\frac{1}{2}} c'(x^*(\Pr(\alpha|\text{piv};\sigma_n^*, n), 1)) = 0. \tag{2.148}$$

Recall (2.51),

$$q(\alpha;\sigma_n^*) - q(\beta;\sigma_n^*) = 2\Big[\int_{t\in\ell} x(t)dH(t) - \int_{t\in s} x(t)dH(t)\Big]. \tag{2.149}$$

Recall (2.115) and (2.44), which imply $\int_{t\in g} x(t)dH(t) \approx c_4 x^*(\hat{p},1)^2 W(g)$ for some constant $c_4 \neq 0$. Hence,

$$x^*(\hat{p},1)^2 \approx c_5\Big[q(\alpha;\sigma_n^*) - q(\beta;\sigma_n^*)\Big]$$

for some constant $c_5 \neq 0$. Then,

$$\begin{aligned} x^*(\hat{p},1)^2 &\leq 2c_5\Big[|q(\alpha;\sigma_n^*) - \frac{1}{2}| + |q(\beta;\sigma_n^*) - \frac{1}{2}|\Big] \\ &\leq 4c_5\frac{\delta_n}{n^{-\frac{1}{2}}}, \end{aligned} \tag{2.150}$$

where I used the triangle equality on the first inequality and $\alpha = \arg\max(|q(\alpha;\sigma_n^*) - \frac{1}{2}|, |q(\alpha;\sigma_n^*) - \frac{1}{2}|)$ for the second inequality. Hence, (2.148) implies

$$\lim_{n\to\infty} nx^*(\Pr(\alpha|\text{piv};\sigma_n^*,n),1)^2 c'(x^*(\Pr(\alpha|\text{piv};\sigma_n^*,n),1)) = 0 \tag{2.151}$$

Using (2.1),

$$\lim_{n\to 0} \frac{x^2 c'(x)}{xc(x)} = d. \tag{2.152}$$

Recall (2.27), hence $x^*(\Pr(\alpha|\text{piv};\sigma_n^*,n),1) \to 0$ as $n \to \infty$. Thus, combining (2.151) and (2.152),

$$\lim_{n\to\infty} nx^*(\Pr(\alpha|\text{piv};\sigma_n^*,n),1)c(x^*(\Pr(\alpha|\text{piv};\sigma_n^*,n),1)) = 0. \tag{2.153}$$

We claim that any equilibrium sequence $(\sigma_n^*)_{n\in\mathbb{N}}$ satisfies

$$\int_{t\in g} c(x(t))dH(t) \approx \frac{2(d-1)}{d}\frac{c(x^*(p^*,1))x^*(p^*,1)}{\chi'(p^*)}W(g). \tag{2.154}$$

The proof follows from previous arguments: the proof is a verbatim to the calculations in section 2.4.2.2, except that we need to replace $x(\Pr(\alpha|\text{piv};\sigma_n^*,n),k)$ with $c(x(\Pr(\alpha|\text{piv};\sigma_n^*,n),k))$ at the appropriate places. Then, (2.153) and (2.154) imply

$$\lim_{n\to\infty}(2n+1)\Big[\int_{t\in[-1,1]^2} c(x(t))dH(t)\Big] = 0. \tag{2.155}$$

Finally, the lemma follows from the weak law of large numbers.

References

Ahn, David S, and Santiago Oliveros. 2012. "Combinatorial voting." *Econometrica* 80 (1): 89–141. [77]

Alesina, Alberto, and Dani Rodrik. 1994. "Distributive politics and economic growth." *quarterly journal of economics* 109 (2): 465–490. [89]

Ali, S Nageeb, Maximilian Mihm, and Lucas Siga. 2018. "Adverse Selection in Distributive Politics." Working paper. Working paper, Penn State University. [89]

Austen-Smith, David, and Jeffrey S Banks. 1996. "Information aggregation, rationality, and the Condorcet jury theorem." *American political science review* 90 (1): 34–45. [87]

Bhattacharya, Sourav. 2013a. "Preference monotonicity and information aggregation in elections." *Econometrica* 81 (3): 1229–1247. [61, 87]

Bhattacharya, Sourav. 2013b. "Preference monotonicity and information aggregation in elections." *Econometrica* 81 (3): 1229–1247. [64, 77]

Billingsley, Patrick. 2008. *Probability and measure*. John Wiley & Sons. [91]

Carpini, Michael X Delli, and Scott Keeter. 1996. *What Americans know about politics and why it matters*. Yale University Press. [59]

Converse, Philip E. 1964. "The nature of belief systems in mass publics (1964)." *Critical review* 18 (1-3): 1–74. [59]

Davis, Burgess, and David McDonald. 1995. "An elementary proof of the local central limit theorem." *Journal of Theoretical Probability* 8 (3): 693–702. [76, 96]

Downs, Anthony. 1957. "An economic theory of political action in a democracy." *Journal of political economy* 65 (2): 135–150. [61]

Durrett, R. 1991. "Probability: Theory and Examples, Wadsworth & Brooks/Cole, Pacific Grove." *MR1068527*, [91, 102]

Eguia, Jon X, and Dimitrios Xefteris. 2018. "Implementation by vote-buying mechanisms." [88]

Feddersen, Timothy, and Wolfgang Pesendorfer. 1997. "Voting behavior and information aggregation in elections with private information." *Econometrica:* 1029–1058. [59–61, 89]

Feddersen, Timothy, and Wolfgang Pesendorfer. 1998. "Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting." *American Political science review* 92 (1): 23–35. [87]

Fernandez, Raquel, and Dani Rodrik. 1991. "Resistance to reform: Status quo bias in the presence of individual-specific uncertainty." *American economic review*, 1146–1155. [61, 89]

Gnedenko, Boris Vladimirovich. 1948. "On a local limit theorem of the theory of probability." *Uspekhi Matematicheskikh Nauk* 3 (3): 187–194. [76]

Heese, Carl. 2020. "Information Cost and Utilitarian Welfare in Elections." Working paper. mimeo. [88]

Heese, Carl, and Stephan Lauermann. 2017. "Persuasion and Information Aggregation in Elections." Working paper. Working Paper. [77]

Henderson, Michael. 2014. "Issue publics, campaigns, and political knowledge." *Political Behavior* 36 (3): 631–657. [59]

Hu, Li, and Anqi Li. 2018. "The Politics of Attention." *arXiv preprint arXiv:1810.11449*, [62]

Iyengar, Shanto, Kyu S Hahn, Jon A Krosnick, and John Walker. 2008. "Selective exposure to campaign communication: The role of anticipated agreement and issue public membership." *Journal of Politics* 70 (1): 186–200. [59]

Krishna, Vijay, and John Morgan. 2011. "Overcoming ideological bias in elections." *Journal of Political Economy* 119 (2): 183–211. [61, 88]

Krishna, Vijay, and John Morgan. 2015. "Majority rule and utilitarian welfare." *American Economic Journal: Microeconomics* 7 (4): 339–375. [88]

Krosnick, Jon A. 1990. "Government policy and citizen passion: A study of issue publics in contemporary America." *Political behavior* 12 (1): 59–92. [59]

Lalley, Steven, and E Glen Weyl. 2018. "Nash equilibria for quadratic voting." *Available at SSRN 2488763*, [61, 88]

Martinelli, César. 2006. "Would rational voters acquire costly information?" *Journal of Economic Theory* 129 (1): 225–251. [61, 74, 86, 88]

Martinelli, César. 2007. "Rational ignorance and voting behavior." *International Journal of Game Theory* 35 (3): 315–335. [88]

Matějka, Filip, and Guido Tabellini. 2017. "Electoral competition with rationally inattentive voters." *Available at SSRN 3070204*, [62]

Oliveros, Santiago. 2013a. "Abstention, ideology and information acquisition." *Journal of Economic Theory* 148 (3): 871–902. [88]

Oliveros, Santiago. 2013b. "Aggregation of endogenous information in large elections." [86, 88]

Persson, Torsten, and Guido Tabellini. 1994. "Is inequality harmful for growth?" *American economic review*, 600–621. [89]

Triossi, Matteo. 2013. "Costly information acquisition. Is it better to toss a coin?" *Games and Economic Behavior* 82: 169–191. [88]

Chapter 3

Motivated Information Acquisition in Social Decisions

Joint with Si Chen

3.1 Introduction

The motivated reasoning literature demonstrates that people often trade off the accuracy against the desirability of their beliefs (for a review, see Bénabou and Tirole, 2016). The desirability of beliefs can arise in decisions where benefiting oneself *might* harm others. In these situations, individuals can behave selfishly without a guilty conscience if they believe that the selfish decision harms no others (for a review, see Gino, Norton, and Weber, 2016). In this paper, we analyze how individuals acquire information about the externalities of the decisions that they are about to make.

To shed light on the dynamics of the information acquisition process, we focus on information that unveils the unknown externalities gradually (i.e., *noisy* information). Whereas a piece of perfect one-shot information uncovers the truth immediately, noisy information increases one's belief accuracy bit by bit. Individuals can not only decide whether to start acquiring noisy information but also *when* to stop the inquiry. Compared to perfect information, situations with noisy information offer individuals a higher chance to end up with beliefs more desirable than their initial beliefs, by allowing them to choose when to stop their inquiries strategically.

In many economic decisions with potential externalities, individuals can acquire noisy information to guide their decisions. Examples include medical examinations that help a doctor to decide between treatments with different profits, media consumption before voting on ethically controversial but personally costly policies, or candidate screening and interviewing by discriminatory employers on the labor market. In these decisions, when individuals decide to stop acquiring noisy information

plays an important role in both the decision-making and the resulting welfare outcomes.

This paper makes three main contributions. (i) We *experimentally* show how individuals strategically decide when to stop acquiring noisy information about their options' externalities when an option benefits themselves. (ii) We propose a *theoretical model* that makes testable predictions about individuals' information choices in social decisions. These predictions are consistent with empirical findings, including the noisy information acquisition strategy found in our experiment. (iii) We show both in theory and in our experimental data that strategic information acquisition motivated by selfish interests can *reduce* the negative externalities resulting from the decision. We present these three contributions in detail below.

First, we conduct a laboratory experiment to investigate the acquisition of noisy information empirically. By doing so, we address three challenges that render an investigation of noisy information acquisition in the field, using observational data, difficult. First, individuals' often unknown and heterogeneous prior beliefs can act as a confounding factor; in our laboratory experiment, we fix the prior beliefs of all subjects such that they begin with the same known prior belief. Second, the information history of each individual is usually hard to monitor; our experiment allows us to monitor the entire information history of each subject. Third, the access to information and interpretation of it are often heterogeneous; the information in our experiment has a clear Bayesian interpretation and is costless for all subjects. Besides, we provide the subjects with the Bayesian posterior beliefs after each piece of information to address heterogeneous ability to interpret information rationally.

More specifically, our subjects take part in a modified binary dictator game, in which each dictator has to decide between two options. The dictators know each option's outcome for themselves. In our baseline, the two options pay the dictators themselves equally. In the treatment, in contrast, one option pays the dictators more than the other option. For each dictator, contingent on an unknown binary state, one of the options reduces the payoff of the receiver, while the other does not. Before making the decision, each dictator can acquire as much noisy information as they want about which option harms the receiver. The information is costless. If one option generates a higher payoff for the dictators, they can opt for the extra payoff without a guilty conscience, as long as they believe that this option does not harm others. Whereas when the options pay themselves equally, the dictators do not have this incentive to prefer certain beliefs about the harmful option. Hence, the dictators in the latter case serve as the baseline.

In the laboratory experiment, we find that compared to the baseline, dictators facing a self-benefiting option *exploit* information: when most of the information received up to that point suggests that the self-benefiting option *harms* the receivers, a higher proportion of them *continue* acquiring information; when most of the information received up to that point suggests that the selfish option causes *no harm* to the receivers, a higher proportion of them *stop* acquiring information. How does this

information acquisition strategy arise? Intuitively, having received dominant information suggesting that the self-rewarding option harms the receivers, the dictators become more inclined to forsake the additional payment. In this case, the further information might present supporting evidence for a selfish decision favorably and make them choose the self-benefiting option instead. In contrast, having received dominant information supporting the innocuousness of the self-rewarding option, individuals face the undesirable risk that further information might challenge the previous evidence. This intuition is formalized in our theoretical model.

As the second contribution, we propose a theoretical model that analyzes the acquisition of information to all degrees of noise. It shows that the information acquisition strategy found in our experiment can be optimal. In our model, a Bayesian agent, who values her belief in her righteousness, attempts to persuade herself to behave selfishly by strategically acquiring information. This self-persuasion modeling approach draws on the Bayesian persuasion model (Kamenica and Gentzkow, 2011). In our model, the sender and the receiver of the signal in Bayesian persuasion are the same person, namely the dictator in our experiment. The agent's signal-sender-self first chooses the information to acquire, and the information pins down her posterior belief distribution. Then the agent's signal-receiver-self chooses the option that maximizes her expected utility given the realized posterior belief. The agent's utility consists of two preference components: preferences for material gains (material utility) and preferences for beliefs that her decision does not harm others (belief utility). Intuitively, in decisions with a self-benefiting option, the optimal information acquisition strategy has two properties: first the agent forgoes her self-interests only when she is certain that doing so benefits others; second, when she chooses the self-benefiting option, her marginal gain of belief utility from being more certain about the state is weakly smaller than the downside risk that the realized posterior belief leans against the self-benefiting option. Leveraging techniques from the Bayesian persuasion model of Kamenica and Gentzkow (2011), our model offers tractable tools for analyzing information acquisition. It generates rich testable predictions, including predictions about the welfare consequences of the motivated information acquisition strategy documented in our experiment.

As a third contribution, we theoretically and empirically show results regarding receiver welfare that might not be obvious at first sight. Although one might think that strategic information acquisition motivated by selfish interests must lead to more negative externalities, our model shows that also the reverse can happen: for some agent types, motivated information acquisition *improves* the welfare of the others affected by the decision. Our experimental data provide evidence consistent with this prediction. This counter-intuitive result arises from a moral hazard problem: when disinterested, some agent types acquire only a small amount of information due to, for example, the satisficing behavior (Simon, 1955). The agent's selfish preference for one option over the other can mitigate this moral hazard problem by causing her to choose her least-preferred option only when she is certain that it is

harmless to others. This result implies that delegating information acquisition to a neutral investigator might lower the welfare of the others affected by the decision.

In terms of the empirical literature, this paper contributes insights into how people engage in motivated reasoning. To the best of our knowledge, we are the first to show that individuals strategically decide when to stop acquiring noisy information, even if they interpret information rationally. The existing literature on motivated beliefs has largely focused on biases in processing *exogenous* information and find that people react to exogenous information in a self-serving manner (Eil and Rao, 2011; Mobius, Niederle, Niehaus, and Rosenblat, 2011; Falk and Szech, 2016; Gneezy, Saccardo, Serra-Garcia, and Veldhuizen, 2016; Exley and Kessler, 2018; Zimmermann, forthcoming). In the literature on excusing selfish behavior without involving information, individuals have been found to manipulate their beliefs and avoid being asked for good deeds (Haisley and Weber, 2010; DellaVigna, List, and Malmendier, 2012; Di Tella, Perez-Truglia, Babino, and Sigman, 2015; Andreoni, Rao, and Trachtman, 2017). An early psychology paper of Ditto and Lopez (1992) documents that individuals require less supportive information to reach their preferred conclusion, possibly due to the bias of overreacting to their preferred information. In comparison, the psychology behind our finding is the tradeoff between a more informed vs. a more desirable decision, rather than the fact that information deemed more valid leads to a conclusion faster. Our experiment shows evidence that individuals use strategic information acquisition itself as an instrument for motivated reasoning.

Our empirical investigation of endogenous information choice relates to the empirical studies on the avoidance of perfectly revealing information in social decisions (Dana, Weber, and Kuang, 2007; Feiler, 2014; Grossman, 2014; Golman, Hagmann, and Loewenstein, 2017; Serra-Garcia and Szech, 2019). In contrast to information avoidance, we find that when it comes to noisy information, individuals *seek* further information if the previously received information is predominantly against the innocuousness of their selfish interests. The avoidance of perfect information documented in the previous studies importantly reveals that individuals have information preferences in social decisions. Delving into *how* people acquire information, our investigation sheds light on *what* the individuals' information preferences are in social decisions. Our model provides a unified framework for analyzing the acquisition of information, with the avoidance of perfect information as a special case.

Another related strand of the empirical literature is the one focusing on rational inattention, showing that individuals who allocate *costly* attention rationally might make decisions based on incomplete information (e.g. Bartoš, Bauer, Chytilová, and Matějka, 2016; Ambuehl, 2017; Masatlioglu, Orhun, and Raymond, 2017). As pointed out by Bénabou and Tirole (2016), when the nature of the decision so determines that some beliefs are more desirable than others, the decision-makers might engage in motivated reasoning and lean towards these beliefs. This is a different psychology than the undirected inattention. For inattention to be rational, information must be costly. In contrast, in our experiment, information entails no monetary

cost and a highly limited time cost. We also limit the cognitive cost to interpret the information by providing Bayesian posterior beliefs to subjects after each piece of information.

In terms of the theory literature, featuring an agent who cares about her own belief that her decision harms no others, our model relates to the literature on belief-dependent utility. Deviating from the outcome-based utility, economic research has put forward concepts of utility directly derived from beliefs, including the utility derived from memories (remembered utility, Kahneman, Wakker, and Sarin, 1997; Kahneman, 2003, etc), the anticipation of future events (anticipatory utility, Loewenstein, 1987; Brunnermeier and Parker, 2005; Brunnermeier, Gollier, and Parker, 2007; Schweizer and Szech, 2018, etc), ego-relevant beliefs (ego utility, Köszegi, 2006, etc), and belief-dependent emotions (Geanakoplos, Pearce, and Stacchetti, 1989, etc). We suggest that individuals receive utility from believing that their decisions impose no harm on others. This approach is most similar to the belief utility from a moral self-identity proposed by Bénabou and Tirole (2011) in the self-signalling games.

By modelling social decisions as driven by utility based on beliefs in one's righteousness, we add to the discussion of an important yet less-understood aspect of social preference, namely social preference under uncertainty. In social decisions with uncertainty, an expected-utility-maximizing agent with intrinsic valuation for the welfare *outcome* of others always prefers complete knowledge in social decisions (for example, the agents in Andreoni, 1990; Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). It contradicts our empirical finding of strategic information acquisition and the avoidance of perfect information observed by, for example, Dana, Weber, and Kuang (2007). Reassessing individuals' motives in social decisions, some models deviate from outcome-based social preferences. Andreoni and Bernheim (2009) propose that individuals act fairly to signal to others that they are fair. Niehaus (2014) proposes a model with an agent who receives a warm glow from her *perceived* social outcomes of her decision. Rabin (1994), Konow (2000), and Spiekermann and Weiss (2016) suggest cognitive dissonance to be a factor for prosocial decisions. In these models, the conflicting desires for selfish interests and fairness create an unpleasant tension, which the agents can reduce by deceiving themselves that a selfish option is fair. A model proposed by Rabin (1995) views moral dispositions as "internal constraints on the agent's true goal of pursuing her self-interest." It shows that for an agent who only engages in a self-benefiting action if she is certain enough that this action harms no one else, partial information or information avoidance can be optimal. In comparison to these studies, our modeling approach connects to the literature of belief-based utility and Bayesian persuasion (Kamenica and Gentzkow, 2011) by modeling an agent who gains utility directly from her beliefs and attempts to persuade herself to behave selfishly. Mathematically, our model includes the agent in Rabin (1995) as a special type.

Another strand of the literature proposes self-signaling as the main concern in social decisions (Akerlof and Kranton, 2000; Bodner and Prelec, 2003; Bénabou and Tirole, 2006; Bénabou and Tirole, 2011; Grossman and Weele, 2017). Assuming a high level of individual rationality, a self-signaling model features intrapersonal signaling games in which one self of the agent knows her prosocial type and makes decisions, including the decision on what information to collect, and the other self observes the decisions to infer her prosocial type. Addressing *whether* people acquire perfect information, Grossman and Weele (2017) endogenize the decision to avoid perfectly revealing information and show that the avoidance of *perfect* information can be an equilibrium outcome in a self-signaling model. In contrast, we model the process of acquiring information as the process of a person persuading herself to behave selfishly. Leveraging insights from the Bayesian persuasion, our model is tractable. It goes beyond the binary decision of acquiring or avoiding a certain type of information and characterizes the optimal information acquisition strategies regarding a large range of information environments.

We organize the rest of the paper as follows: In Section 3.2, we first detail the experimental design and then empirically analyze the dictators' information acquisition strategy in our experiment. In Section 3.3, we present the theoretical model that predicts our empirical findings. In Section 3.4, we theoretically show that strategic information acquisition motivated by the dictator's selfish interests can improve the receiver welfare. We also provide consistent results in our experimental data. In Section 3.5, we conclude and propose some ideas for future research.

3.2 Motivated Information Acquisition

This section focuses on how individuals acquire information about their options' externalities in a decision. In Section 3.2.1, we provide details of the experimental design. In Section 3.2.2, we empirically analyze the dictators' information acquisition strategies.

3.2.1 A Laboratory Experiment With Modified Dictator Games

We conduct a laboratory experiment with modified binary dictator games. Contingent on an unknown state, one of the two options of the dictator game reduces the receivers' payoffs, and the other does not. Before deciding, the dictators can acquire information about the harmful option at no cost.

3.2.1.1 The Treatment Variations

Our experiment has a 2×2 design and 4 treatments, as illustrated in Table 3.1. The treatments vary on two dimensions: (i) whether one of the dictator game options

increases the dictators' payoffs; (ii) whether the dictators can proceed to the dictator game without acquiring any information on the externalities of their options.

The key treatment variation in our experiment is whether the dictators' selfish interests are concerned in the dictator game. In the "*Tradeoff*" treatments, one option increases the dictators' payoffs, while the other does not. In the "*Control*" treatments, neither option affects the dictators' payoffs. The comparison between the *Tradeoff* and *Control* pins down the causal effect of having a self-benefiting option on the dictators' information acquisition behavior. We describe the details of this treatment variation below when we present the dictator game.

The second treatment variation concerns the dictators' freedom to acquire no information. It serves two purposes: (i) In the "*NoForce*" treatments, dictators are *not forced* to acquire any information. These treatments allow us to examine the proportion of dictators who do not acquire any information, but they also leave room for self-selection into the information processes. (ii) In the "*Force*" treatments, the dictators are *forced* to acquire at least one piece of information before making their decisions in the dictator game. This modification eliminates the potential self-selection into the information process.[1]

Table 3.1. Treatments

	With Selfish Interests	No Selfish Interests	Shorthand
No Forced Draw	*Tradeoff–NoForce*	*Control–NoForce*	***NoForce***
A Forced Draw	*Tradeoff–Force*	*Control–Force*	***Force***
Shorthand	***Tradeoff***	***Control***	-

This table presents our four treatments with a two by two design. *Tradeoff* vs. *Control* is our key treatment variation. Dictators in *Tradeoff* can gain additional payment by choosing a particular option in the modified dictator game, while those in *Control* cannot. *Force* vs. *NoForce* differ in that in the former the dictators have to acquire at least one piece of information, while in the latter they can choose to acquire no information.

3.2.1.2 The Dictator Game

Table 3.2 presents the payment scheme of the dictator game in the *Tradeoff* and the *Control* treatments respectively. In all treatments, the dictators choose between two options, x and y. There are two states of the world, *"x harmless"* or *"y harmless"*. Depending on the state, either option x or y reduces the receivers' payments by 80 points, while the other one does not affect the receivers' payment. Note that each option harms the receiver in one of the states. This design makes sure that the dictators cannot avoid the risk of harming the receiver without learning the state. In *Control*, the dictators receive no additional points regardless of their choices and

1. We explain in details this selection effect when we analyze the data in Section 3.2.2.

Table 3.2. Dictator Decision Payment Schemes

(a) *Control* Treatments

	Good state (x harmless)	Bad state (y harmless)
x	$(\mathbf{0}, 0)$	$(\mathbf{0}, -80)$
y	$(0, -80)$	$(0, 0)$

(b) *Tradeoff* Treatments

	Good state (x harmless)	Bad state (y harmless)
x	$(+\mathbf{25}, 0)$	$(+\mathbf{25}, -80)$
y	$(0, -80)$	$(0, 0)$

These tables present the dictator games in *Control* and *Tradeoff* treatments. The number pairs in the table present (dictator's payment, receiver's payment).

the state. In *Tradeoff*, x is self-benefiting for the dictators: they receive 25 additional points when choosing x, but no additional points when choosing y.

***Good* State vs *Bad* State.** For the ease of exposition, we hereafter refer to the state *"x harmless"* as the *"Good state"*, and the state *"y harmless"* as the *"Bad state"*. It is because in state *x harmless*, the dictator's and the receiver's interests are aligned in *Tradeoff*: option x is better for both of them. The dictator can claim the additional payment of 25 points without harming the receiver. Reversely, in state *y harmless*, if the dictator decides to choose x to gain the additional payment, she makes the receiver worse-off. The dictator is in a dilemma between less payment for herself or hurting the receiver. Although this contrast between states does not apply to the *Control* treatments, we will refer to *"x harmless"* as the *Good state* and *"y harmless"* as the *Bad state* for consistency.

Note that in treatments *Tradeoff*, dictators would prefer to believe that they are in the *Good state*, such that they can choose option x and gain the additional payment without having a bad conscious; whereas in the *Control* treatments, dictators are indifferent about which state they are in, since their payments are not affected by their decisions in either state.

The dictators start the experiment without knowing the state that they are in individually. They only know that in every twenty dictators, seven are in the *Good state*, and thirteen are in the *Bad state*. That is, the dictators start the experiment with a prior belief of 35% on that they are in the *Good state* and 65% in the *Bad state*. Before making the decision, they can update their beliefs by drawing information described in the next subsection.

3.2.1.3 The Noisy Information

We design a noisy information generator for each state, which generates information that is easily interpretable according to the Bayes' rule. Specifically, each piece of information is a draw from a computerized box containing 100 balls. In *Good state*, 60 of the balls are white and 40 are black; in *Bad state*, 40 balls are white and 60 are black (Figure 3.1). The draws are with replacement from the box that matches to

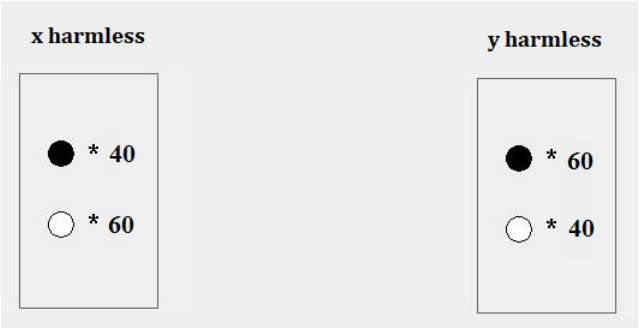


Figure 3.1. The Noisy Information Generators

each dictator's actual state. After each draw, we display the Bayesian posterior belief on the individual computer screen, to reduce the cognitive cost of interpreting the information and reduce non-Bayesian updating.

Good News vs. Bad News. For the ease of exposition, we refer to a white ball as a piece of "good news" and a black ball as a piece of "bad news". It is because, in the *Good state*, dictators draw a white ball with a higher probability. A white ball hence supports the dictators to believe in the *Good* state, in which the dictators in treatments *Tradeoff* can choose x and gain the additional payment without reducing the payment of the receiver. Reversely, in the *Bad state*, dictators would draw a black ball with higher probability. A black ball is an evidence for the *Bad* state, in which option x rewards the dictators in *Tradeoff* at the cost of the receivers. Although dictators in *Control* do not have a preference over the two states, and hence unlikely to have a preference for black or white balls, we will still refer to a white ball as good news and a black ball as bad news for consistency.

3.2.1.4 The Experimental Procedure

The experiment consists of three parts: the preparation stage, the main stage, and the supplementary stage.

The Preparation Stage:. (i) The dictators read paper-based instructions on the dictator decision, and the noisy information. (ii) We also describe in written the Bayes rule and tell the dictators that later in the experiment, we are going to help them to interpret the information by showing them the Bayesian posterior beliefs after each ball that they draw. (iii) Besides, the instructions specify that each experiment participant starts the experiment with 100 points of an endowment. (iv) We also inform them that option x is harmless for 7 out of 20 of the dictators and y for 13 out of 20. That is, the dictators' prior beliefs on the states are 35% and 65% on the *Good* state and the *Bad* state.

After reading the instructions, the dictators answer five control questions designed to check their understanding of the instructions. They keep the paper instructions for reference throughout the experiment.

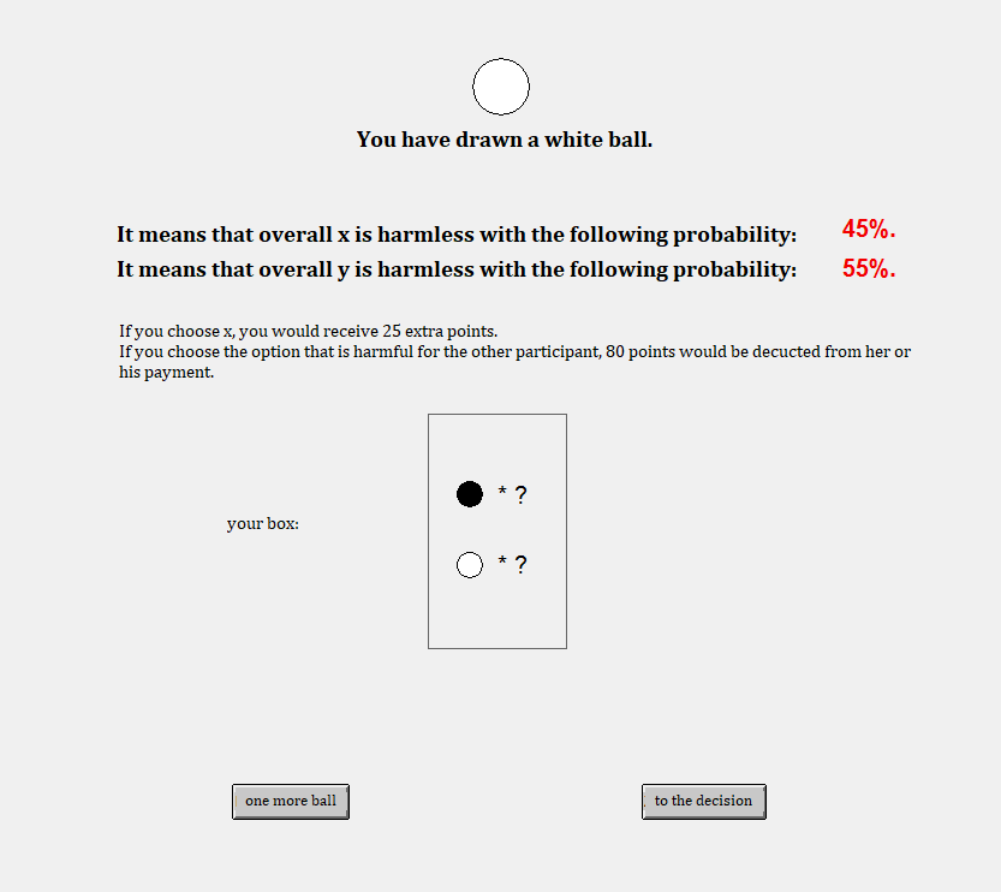


Figure 3.2. Screenshot of the Information Stage

The Main Stage:. In the main stage, (i) dictators can acquire information about the state that they are individually in; (ii) they choose between x and y in the dictator game.

Specifically, the dictators can acquire a piece of information by clicking a button that makes the computer draw a ball randomly from the box matched to their actual individual state (see Figure 3.1). The draws are with replacement. After each draw, the screen displays the latest ball drawn and the Bayesian posterior beliefs on the *Good state* and the *Bad state* given all the balls drawn so far (rounded to the second decimal, see Figure 3.2). There are two buttons on the screen: one to draw an additional ball, and the other to stop drawing and proceed to the dictator game. Either to draw a ball or to stop drawing, a dictator must click on one of the buttons.

The draws do not impose any monetary cost on the dictators. The time cost of acquiring information is limited: between draws, there is a mere 0.3 second time

lag to allow the ball and the Bayesian posterior belief to appear on the computer screen. It means that a dictator can acquire 100 balls within 30 seconds, which would almost surely yield certainty.

In the *NoForce* treatments, the dictators can draw from zero to infinitely many balls. That is, they can proceed directly to the dictator game without drawing any ball, and if they decide to acquire information, the information acquisition can only be ended by them. In the *Force* treatments, the dictators must draw at least one ball, and after the first draw, they have full autonomy regarding when to stop drawing just like in *NoForce*. Besides drawing balls, the dictators have no other way to learn about the true state that they are in throughout the experiment. It is common knowledge that the receivers do not learn the information acquired by the dictators throughout the experiment.

Having ended information acquisition, dictators choose between x and y in the dictator game in Table 3.1a (in the *Control* treatments) or Table 3.1b (in the *Tradeoff* treatmentss). Next in the implementation state, the dictator' choices are implemented and the payments are calculated.

The Supplementary Stage:. (i) We elicit the dictators' posterior beliefs on the state after the dictator game. The belief elicitation is incentivized by using the randomized Quadratic Scoring Rule. We compare the elicited and the Bayesian posterior beliefs in Appendix 3.A.5 and find that for the majority of dictators, their elicited posterior beliefs and their Bayesian posterior beliefs coincide. (ii) The subjects take part in the Social Value Orientation (SVO) slider measure, which measures "the magnitude of concern people have for others' and categorizes subjects into altruists, prosocials, individualists, and competitive type (Murphy, Ackermann, and Handgraaf, 2011). (iii) The subjects answer a questionnaire consisting of socio-demographics, preferences, a selection of HEXACO personality inventory (Lee and Ashton, 2018), and a 5-item Raven's progressive matrices test (Raven et al., 1998). We report the details of the questionnaire in Appendix 3.A.5.

Implementation:. We randomize within each laboratory session: (i) the *Tradeoff* and *Control* treatments, (ii) the states: we randomly assign 35% of the laboratory terminals to the *Good state*, and 65% to the *Bad state*. The subjects are then randomly seated and randomly matched in a ring for the dictator game. The subjects are told that their decisions would affect the payment of a random participant in the same experimental session other than themselves. After all the subjects have decided in the dictator game, the experiment moves on to the implementation stage, where we inform the subjects that the dictator game decisions are being implemented and their payments are affected according to another participant's dictator game decision. Each subject plays the dictator game only once.

We conducted the experiment in October and December 2018 at the BonnEconLab (*NoForce* and *Force* treatments respectively). 496 subjects took part (168 in *Tradeoff–NoForce*, 167 in *Control–NoForce*, 82 in *Tradeoff-Force* and 79 in *Control-*

Force). Among the subjects, 60% are women, and 93% are students. They are, on average, 24 years old, the youngest being 16 and the oldest being 69. The subjects are balanced between treatments, concerning gender, student status, and age (see Appendix 3.A.5). We used z-tree (Fischbacher, 2007) to implement the experiment and hroot (Bock, Baetge, and Nicklisch, 2014) to invite subjects and to record their participation. Instructions and interfaces on the client computers were written in German, as all subjects were native German speakers.

Payments:. In the experiment, payments are denoted in points. One point equals 0.05 EUR. At the end of the experiment, the details of the points and the equivalent payments earned in the experiment are displayed on the individual computer screens. The subjects received payments in cash before leaving the laboratory. The total earnings of a subject were the sum of the following components: an endowment of 5 EUR, an additional 1.25 EUR if the subject was in treatments *Tradeoff* and chose x, a 4 EUR reduction if the subject's randomly assigned dictator made a decision that reduces her payments, a random payment of either 1.5 EUR or 0 for revealing their posterior beliefs, a payment ranging from 1 to 2 EUR depending on the subject's decisions in the SVO slider measure, a payment ranging from 0.3 to 2 EUR depending on the decisions in the SVO slider measure of another random subject in the same laboratory session, and a fixed payment of 3 EUR for answering the questionnaire. A laboratory session lasted, on average, 45 minutes, with an average payment of 11.14 EUR.

3.2.2 Empirical Analyses of Motivated Information Acquisition

In this section, we analyze the data from our experiment to investigate the effect of having a selfishly preferred option on how individuals acquire information about their options' externalities. The median number of balls drawn by the dictators is 6 (*Tradeoff*: 6, *Control*: 5; Mann-Whitney-U $p = 0.98$). We summarize our data in Appendix 3.A.1 and proceed below with the analyses of the dictators' information acquisition behavior.

Consequently, the effect of our interest – the effect of the information histories on acquiring further information – might be confounded by the prior beliefs. These two effects can be disentangled by comparing treatments *Tradeoff* to Treatments *Control* since the prior beliefs are the same across treatments. The latter treatments serve as a baseline.

Do dictators acquire information?

Finding 1. The proportion of dictators who do not acquire any information is 15% in *Tradeoff–NoForce* and 7% in *Control–NoForce*.

In the *NoForce* treatments, where the dictators are allowed to draw no information before the dictator game, 38 out of 335 proceed to the dictator game without

drawing information (*Tradeoff–NoForce:* 15%; *Control–NoForce:* 7%). Among them, in *Tradeoff–NoForce*, 25 out of 26 choose x, the option with additional payments for themselves; in *Control– NoForce*, where neither option produces additional payments for the dictators themselves, only 2 out of 12 choose x.

Table 3.3. Proportion of Dictators Drawing No Ball

	No Info%	Their Choices	
Tradeoff–NoForce	15%	x: 96%	y: 4%
Control–NoForce	7%	x: 17%	y: 83%
Chi-2 p-value	0.02	0.00	

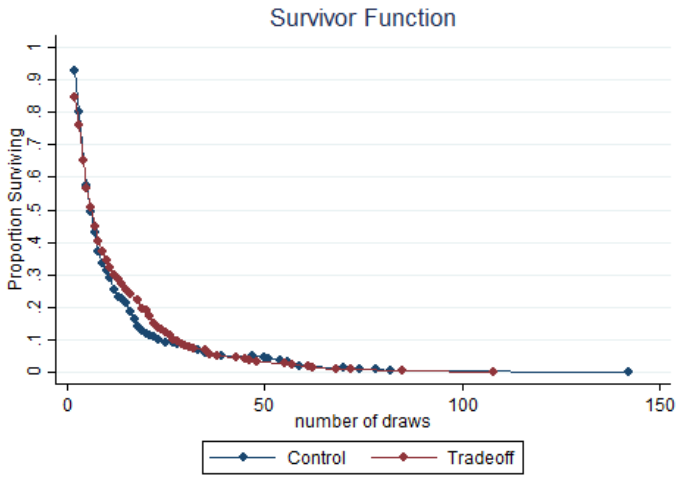
This table displays in each treatment (i) the proportion of dictators who do not draw any ball before making their decisions between x and y; (ii) the proportion *among them* who choose option x. Note that in treatment *Tradeoff–NoForce*, dictators who choose option x receive additional payment, while those in treatment *Control–NoForce* do not.

Do dictators stop earlier in *Tradeoff* than in *Control*?

Finding 2. Overall, the proportions of dictators who continue acquiring information after each draw do not differ between treatments.

Figure 3.3 presents in *Tradeoff* and *Control* the proportions of dictators surviving over time, i.e. the proportion of dictators who are still acquiring information over time. The survival function does not differ between *Tradeoff* and *Control* (log-rank test for equality of survivor functions, $p = .63$).

Finding 2 speaks against an overall lower propensity to acquire noisy information when individuals' selfish interests are involved in the decision. It contrasts the avoidance of perfect information found by the previous literature (e.g. Dana, Weber, and Kuang, 2007).



This figure plots the fraction of dictators remaining in the information acquisition process over the number of draws.

Figure 3.3. Life Table Survival Function

When do dictators stop acquiring information? We now turn to the 458 dictators who did acquire information and focus on the role of the information history in their decisions to continue acquiring information after each draw of ball.

Specifically, we predict:

Having an option that generates additional payoffs for the dictators themselves (i) increases their tendency to *continue* acquiring information, when a dominant amount of information received up to that point is *bad* news against the innocuousness of this option; (ii) but increases their tendency to *stop*, when a dominant amount of information received is *good* news supporting the innocuousness of this option.

The intuition of the prediction is that when the dictators are inclined to forgo their selfish interests upon receiving dominant bad news, continuing the inquiry might reverse the previous bad news favorably and make them choose the self-rewarding x instead. This possibility might encourage dictators to continue drawing balls. However, when the dictators have received dominant desirable good news and are inclined to behave selfishly, the further information might be bad news that deteriorates their current desirable beliefs. This risk might discourage the dictators from drawing further information. This intuition is formalized in the theoretical model presented in Section 3.3.

In what follows, we first compare the decisions to stop acquiring the information directly after the first draw between *Tradeoff* and *Control*. Then, we analyze the entire information histories, levering insights from the research of survival analysis.

3.2.2.1 The First Draw of Ball

For dictators, whose first ball is good news and those whose is bad news, we respectively compare between *Tradeoff* and *Control* their decisions to continue acquiring information right after the first draw. The good and bad nature of the first draw is exogenous in our experiment since the composition of the 100 balls in the boxes depends solely on the exogenous state, and the draws are random.

Finding 3. (i) When the first draw is bad news, the proportion of dictators who continue drawing balls right after it is similar across treatments. (ii) In case of good news, the proportion is smaller in *Tradeoff* treatments than in *Control* treatments.

Finding 3 shows evidence that having a self-rewarding option causes individuals to be more likely to stop acquiring further information when the previous information supports the innocuousness of this option. On the opposite, when the information received up to that point suggests that the selfish decision harms others, individuals continue acquiring information similarly with or without the self-rewarding option. Table 3.4 and Figure 3.4 present the exact proportions of dictators who continue acquiring information right after the first draw.

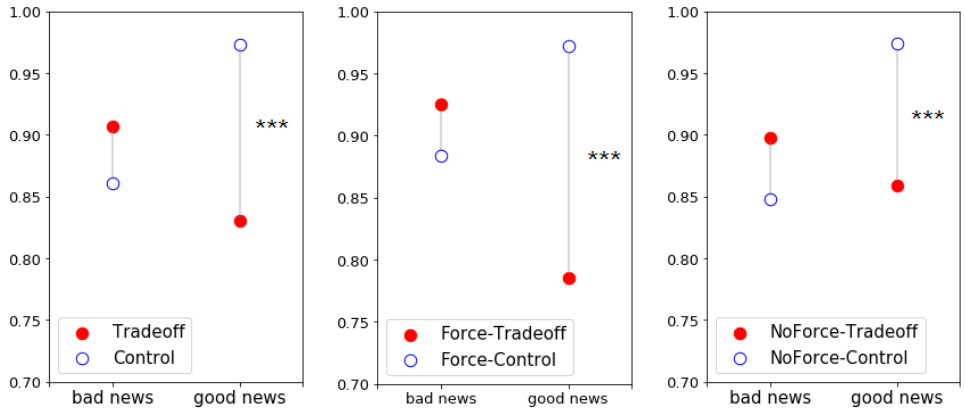
Table 3.4. Proportion of Dictators Continuing After the First Ball

	First News **Good**			First News **Bad**		
Treatment	Pooled	*Force*	*NoForce*	Pooled	*Force*	*NoForce*
Tradeoff	83%	79%	86%	91%	93%	90%
Control	97%	97%	97%	86%	88%	85%
Chi-2 p-value	.00	.01	.00	.26	.52	.22

This table displays the proportions of dictators who continue acquiring information after the first draw in the respective treatment, given the respective first draw. In the *Force* treatments, dictators have to draw at least one ball before choosing between x and y. Note that in the *Control* treatments the within treatment differences given different news are due to the asymmetric prior belief of 35% in the *Good* state.

Discussion. Finding 3 is less prominent in the *NoForce* treatments, where the dictators can choose to draw no information. The reason might be the fact that the dictators in *NoForce* have selected themselves into the information process.

In *Tradeoff–NoForce*, almost all dictators who do not acquire information choose x directly. Had they received a further piece of good news, they would also be willing to stop immediately. Therefore, the *Tradeoff–NoForce* dictators' sorting out of the information process decreases the proportion of them who stop directly after the first good news and reduces the observed effect of the treatment. Similarly, in treatment *Control–Force*, almost all dictators who do not acquire information choose y directly. Had they received a piece of bad news, they might also stop immediately to choose y. Therefore, the self-selection out of the information process of the *Control–Force*



(a) Pooled Data **(b)** *Force* Treatments **(c)** *NoForce* Treatments

These figures present the proportion of dictators who continue acquiring information after the first draw.

Figure 3.4. Proportion of Dictators Continuing after the First Draw

dictators decreases the observed proportion of them who stop right after the first bad news and hence is also against our finding.

3.2.2.2 The Entire Information Histories

Now we turn to the dictators' complete information acquisition process. Each dictator's information history evolves over time. To be able to include it in our analyses, we first split each dictator' complete information history at the unit of one draw.[2] The resulting data set consists of records at the person-draw level. For every draw of each dictator, the pseudo-observation records the dictator's information history up to that draw, whether the dictator chooses to stop or continue acquiring the information directly after that draw and time-constant characteristics of the dictator such as her identity, treatment assignment, and gender. After each draw, we can distinguish between information histories dominated in amount by good and bad news, using a binary dummy variable.

In the framework of a Cox proportional hazard model, we compare the decision to stop acquiring further information between treatments, given these two types of information histories: one dominated in amount by good news, and the other by bad news.[3]

2. Time-varying covariates in survival analysis are often obtained by the method of splitting episodes (see Blossfeld, Rohwer, and Schneider, 2019, pp 137-152).

3. The Cox model has the advantage that the coefficient estimates are easy to interpret. We report a robustness check using the logistic model in Appendix 3.A.4. The results of the logistic model are in line with those of the Cox model.

We are interested in the dictators' hazard to stop acquiring information. The Cox proportional hazard model factors the hazard rate to stop acquiring information into a baseline hazard function $h_0(t)$ and covariates X_t that shift the baseline hazard proportionally, as in (3.1). The baseline hazard function $h_0(t)$ fully captures the time dependency of the hazard.[4]

$$h(t|X_t) = h_0(t) \cdot exp(X_t \cdot b). \tag{3.1}$$

Our model specification is as follows:

$$h(t|X) = h_0(t) \cdot exp(\beta_1 \text{Tradeoff} + \beta_2 \text{Info} + \beta_{12} \text{Tradeoff} \times \text{Info} + \alpha z_t), \tag{3.2}$$

where "Tradeoff" is a dummy variable for treatment *Tradeoff*, "Info" is a categorical variable denoting information histories that are dominated by bad news, good news, or balanced between the two, with bad news dominance as the baseline. z_t is a control variable that measures the accuracy of the individual belief after each ball drawn.[5] After controlling for the belief accuracy, the color of the balls per se appears to have no significant effect on dictators' stopping decisions, as shown later in Table 3.5. To allow for different shapes of the hazard function with respect to gender, cognitive ability (measured by Raven's matrices test) and prosocial types (categorized by SVO measure by Murphy, Ackermann, and Handgraaf, 2011), we stratify the Cox model by these variables (Allison, 2002).[6]

We are interested in the following two hazard ratios:
(i) the first one reflects the effect of the treatment on the hazard rate, given *bad* news dominance in the information history. That is, *ceteris paribus*

$$\begin{aligned} \text{HR}_{Bad} = \frac{h(t|\text{Bad, Tradeoff } = 1)}{h(t|\text{Bad, Tradeoff} = 0)} &= \frac{exp(\beta_1 \cdot 1 + \beta_2 \cdot 0 + \beta_{12} \cdot 1 \cdot 0 + \alpha z_t)}{exp(\beta_1 \cdot 0 + \beta_2 \cdot 0 + \beta_{12} \cdot 0 \cdot 0 + \alpha z_t)} \\ &= \frac{exp(\beta_1 + \alpha z_t)}{exp(\alpha z_t)} \\ &= exp(\beta_1); \end{aligned} \tag{3.3}$$

(ii) the second one reflects the effect of the treatment on the hazard rate, given *good* news dominance in the information history. That is, *ceteris paribus*

4. Unlike many other regression models, the Cox model naturally includes no constant term, since the baseline hazard function already captures the hazard rate at covariate vector 0 (see for example Cleves, Gould, Gutierrez, and Marchenko, 2010).

5. We use the following score as a proxy for the accuracy of beliefs: $\text{belief}_{\text{Good}} \times \text{belief}^2_{\text{Bad}} + \text{belief}_{\text{Bad}} \times \text{belief}^2_{\text{Good}}$. It is a probabilistic belief's expectated Brier score (Brier, 1950). Brier score is a proper score function that measures the accuracy of probabilistic predictions.

6. As shown in Table 3.5, after the stratification, our main covariates affect the hazard to stop acquiring information proportionally. That is, the proportional hazard assumption of the Cox model is not violated.

$$\begin{aligned}\text{HR}_{Good} = \frac{h(t|\text{Good, Tradeoff } = 1)}{h(t|\text{Good, Tradeoff} = 0)} &= \frac{exp(\beta_1 \cdot 1 + \beta_{2,\text{Good}} \cdot 1 + \beta_{12,\text{Good}} \cdot 1 \cdot 1 + \alpha z_t)}{exp(\beta_1 \cdot 0 + \beta_{2,\text{Good}} \cdot 1 + \beta_{12,\text{Good}} \cdot 0 \cdot 1 + \alpha z_t)} \\ &= \frac{exp(\beta_1 + \beta_{2,\text{Good}} + \beta_{12,\text{Good}} + \alpha z_t)}{exp(\beta_{2,\text{Good}} + \alpha z_t)} \\ &= exp(\beta_1 + \beta_{12,\text{Good}}). \qquad (3.4)\end{aligned}$$

Our prediction suggests that HR_{Bad} is smaller than 1 and HR_{Good} is larger than 1. That is, (i) $\beta_1 < 0$; (ii) $\beta_1 + \beta_{12,\text{Good}} > 0$.

In Table 3.5, we report the Cox model results, with standard errors clustered at the individual level. Pooling all treatments, the Cox model coefficient estimates yields Finding 4.

Finding 4. (i) Having received more bad news than good news, the dictators are more likely to **continue** acquiring information in *Tradeoff* than in *Control*; (ii) while they are more likely to **stop** in *Tradeoff* than in *Control*, having received more good news than bad news.

The estimated coefficient of the treatment dummy is $\beta_1 = -.28$, and its interaction with the categorical variable indicating good news dominance is $\beta_{12} = .43$, both significant at 5 percent level. If bad news dominates the information history, the hazard to stop acquiring information in *Tradeoff* is $exp(-.28) = .76$ of that in *Control*, i.e. 24% lower in *Tradeoff*. In contrast, if good news dominates, the hazard in *Tradeoff* is $exp(-.28 + .43) = 1.16$ of that in *Control*, i.e. 16% higher in *Tradeoff*. That is, the treatment of having a selfishly preferred option makes the dictators more likely to continue acquiring information, when they have received predominantly bad news, and more likely to stop when they have predominantly good news. The estimation in the *Force* and *NoForce* treatments point in the same direction.

The Role of Cognitive Ability. When we focus exclusively on dictators above the median cognitive ability, as measured by Raven's matrices test (Table 3.6), we find that the effects in Finding 4 become *stronger* than the average effect that we report in Table 3.5. Having received more bad news, these dictators' hazard to stop acquiring information in *Tradeoff* is $exp(-.35) = .70$ of that in *Control*. Having received more good news, the hazard to stop acquiring information in *Tradeoff* is $exp(-.35 + .59) = 1.30$ of that in control. In comparison, considering all dictators, these numbers are .76 and 1.16, indicating that the tendency to acquire information strategically is more moderate averaging across dictators with all levels of cognitive ability than focusing on the ones with high cognitive ability. This finding suggests that the information acquisition behavior in Finding 4 is more likely out of strategic considerations than due to limited cognitive abilities.

Table 3.5. The Cox Proportional Hazard Model Results

		Pooling All		Force	NoForce
$\hat{\beta}_1$	treatment *Tradeoff*	-.28**	-.24*	-.38*	-.18
		(.12)	(.13)	(.21)	(.16)
$\hat{\beta}_{12}$	Tradeoff ×				
	Good news dominance	.43**	.41**	.32	.42*
		(.20)	(.21)	(.39)	(.26)
	Balanced	-.35	-.42	-.59	-.34
		(.38)	(.38)	(.69)	(.47)
$\hat{\beta}_2$	Good news dominance	-.14	-.23	-.18	-.23
		(.16)	(.12)	(.31)	(.22)
	Balanced	-.52**	-.56**	-.48	-59**
		(.24)	(.22)	(.38)	(.30)
Control Variables:					
Belief Accuracy		Yes	Yes	Yes	Yes
Gender, IQ, Prosociality FEs		Yes	Yes	Yes	Yes
Force treatment FE		No	Yes	–	–
Observations (individuals)		458	458	161	297
Chi2 p-value		.00	.00	.00	.00
Violation of PH		NO	NO	NO	NO

This table presents the estimated *coefficients* of the Cox model in (3.2), with standard errors clustered at the individual level. *, **, and *** denote significance at the 10, 5, and 1 percent level. The dependent variable is the hazard to stop acquiring information, and the key coefficients of interests are $\hat{\beta}_1$ and $\hat{\beta}_{12}$. $\exp(\hat{\beta}_1)$ reflects the treatment effect on the dictators' hazard to stop acquiring further information, given information histories dominated by bad news; and $\exp(\hat{\beta}_1 + \hat{\beta}_{12}|\text{Good news dominance})$ reflects the treatment effect on the hazard, given information histories dominated by good news (derivation see Equation (3.4)).

The fixed effects are taken into account by stratification, which allows the baseline hazard to differ according to the control variables, i.e., gender, the prosocial types (categorized by the SVO test), and the cognitive ability (measured by Raven's matrices test). We also control for the belief accuracy, measured by the Brier score of the beliefs after each draw (see Footnote 5). The reported likelihood Chi-square statistic is calculated by comparing the deviance ($-2\times$ log-likelihood) of each model specification against the model with all covariates dropped. The violation of the proportional hazard assumption of the Cox model (PH) is tested using Schoenfeld residuals. In all four cases, the PH is not violated for each covariate nor globally. We use the Breslow method to handle ties.

Table 3.6. The Cox Model Results For Above and Below Median Raven's Scores

		Above Median	Below Median
$\hat{\beta}_1$	treatment *Tradeoff*	-.35**	-.17
		(.16)	(.20)
$\hat{\beta}_{12}$	Tradeoff ×		
	Good news dominance	.59**	-.21
		(.27)	(.30)
	Balanced	.32	-1.03*
		(.54)	(.59)
$\hat{\beta}_2$	Good news dominance	-.10	-.25
		(.22)	(.27)
	Balanced	-.98**	-.21
		(.40)	(.32)
Control Variables:			
Belief Accuracy		Yes	Yes
Gender, IQ, Prosociality FEs		Yes	Yes
Force treatment FE		No	Yes
Observations (individuals)		267	191
Chi2 p-value		.00	.00
Violation of PH		NO	NO

This table presents the Cox model results for the subjects above and below median cognitive ability, measured by the number of correctly answered questions in Raven's matrices test, pooling data from all treatments. Standard errors are clustered at the individual level. The median number of correct answers to Raven's test is four out of five in our experiment. In this table, the subjects above the median have given correct answers to four or five questions in Raven's test, and the subjects below the median have correctly answered below four questions in Raven's test. We find that subjects with higher cognitive ability have a higher tendency to acquire information strategically. For a comprehensive table description, please see that of Table 3.5.

3.3 Optimal Information Acquisition in Theory

In this section, we present a model that characterizes individuals' optimal information choices in decisions affecting others. Its predictions include the noisy information acquisition strategy found in Section 3.2.2 and the avoidance of perfect information evidenced by Dana, Weber, and Kuang (2007) (see Appendix 3.A.6).

Heavily drawing on the Bayesian persuasion model (Kamenica and Gentzkow, 2011), we assume Bayesian updating and transform the problem of information acquisition to the problem of *self-persuasion*.[7] We compare the optimal information acquisition strategy between two scenarios: a decision in which one of the options benefits the agent herself (like in *Tradeoff*), and a decision in which her benefits are not concerned (like in *Control*).

To make the idea of *self-persuasion* concrete, let us consider the dictators in our experiment and answer two questions. First, which option would the dictators persuade themselves to choose? In *Control*, the payment of a dictator is not affected by her choice between the two options x and y. She hence has no incentive to persuade herself of either of the options. Only in *Tradeoff* where a dictator receives additional payment for choosing x, the dictator has the incentive to persuade herself to choose x.

Second, why would a dictator in *Tradeoff* need self-persuasion at all? If she cares more about her own payment, she can choose x to claim the additional payment, without drawing any information. Or if she cares more about the receiver's payment, she can acquire information until she is sufficiently certain about the state, and decide accordingly. Our observation that the majority of dictators in *Tradeoff* do neither of the above points at a third possibility – while caring for their own payments, the dictators also want to *believe* that they are not harming the receivers. This is where self-persuasion comes into play. Although the problem concerns only one person, i.e., the dictator, we can understand her as having two selves: a *Sender-self* who affects her beliefs by acquiring information, but does not make the decision between x and y; and a *Receiver-self*[8] who has a plan regarding which option to choose given any belief in the states, she is the self making the dictator decision. By acquiring information, the dictator's *Sender-self* sends signals to persuade her *Receiver-self* to choose the self-benefiting option x.

Information can be used by an individual's *Sender-self* to persuade her *Receiver-self* to make a certain decision because it affects her beliefs. For example, in our experiment, information influences a dictator's belief about which state she is in. This belief in turn affects her decision between x and y. A Bayesian agent's beliefs

7. In the experiment, we facilitate Bayesian updating by providing the dictators with the Bayesian posterior beliefs after each draw of information.

8. The *Receiver-self* is to be distinguished from the receiver in the dictator game. The former refers to the receiver of signals in the Bayesian persuasion model (Kamenica and Gentzkow, 2011).

only change when responding to information. When committing to acquiring information in a certain way, the *Sender-self* is committing to the distribution of her posterior beliefs. She can choose any distribution of posterior beliefs that is Bayes-plausible, i.e., any posterior belief distributions with expectation equal to her prior belief.

To illustrate how information persuades, consider a dictator in treatment *Trade-off* whose *Receiver-self* chooses x whenever her belief in the *Good* state is not lower than 40%. Recall that her prior belief in the *Good* state is 35%. Hence without information, this *Receiver-self* would choose y. By acquiring information, the dictator's *Sender-self* can with some probability make her *Receiver-self* choose x instead. For example, if the *Sender-self* sends her *Receiver-self* full information that reveals the state completely, the *Sender-self* has a 35% chance to persuade her *Receiver-self* to choose x. By not acquiring full information, the *Sender-self* can do even better. For example, by choosing an information acquisition strategy that yields either 70% or 0% posterior belief in the *Good* state, she can persuade her *Receiver-self* to choose x 50% of the time, i.e. whenever 70% posterior is realized. By reducing the certainty in the *Good* state for which the *Sender-self* makes the *Receiver-self* choose x, the *Sender-self* can increase the probability of successful persuasion. However, the downside of doing so is that she would be less certain that she is not harming others when choosing x – her belief utility of choosing x falls. The optimal information acquisition strategy is determined by trading-off the probability of successful persuasion and the belief utility of being certain that the decision does not harm others.

Our model analyzes the optimal information acquisition strategy for an agent to persuade herself to choose a selfish option when she also wants to believe that her decision does not harm others. We focus on binary action space and binary state space, like in our experiment. The optimal information acquisition strategies in the model are in line with our empirical findings in the experiment.

In Section 3.3.1, we set up the model. In Section 3.3.2, we compare the optimal information acquisition in two scenarios: one in which the receiver's choice solely affects a third person, one in which a certain option benefits the decision-maker herself. Finally, in Section 3.3.3, we show in our data direct evidence of the model predictions.

3.3.1 Setup of the Model

An agent (she) has to make a decision between two options x and y. There is an unknown binary state $\omega \in \{X, Y\} = \Omega$ and the prior belief is that the probability of X is $p_0 \in (0, 1)$. A passive agent, whom we hereafter refer to as *the other* (he), can be affected by the agent's decision between x and y – when the agent chooses an action that does not match the state, i.e. x in Y or y in X, the action has a negative

externality of -1 on the other (he) and otherwise not.[9] The agent dislikes the belief that her decision harms the other. When the agent believes that state X holds with probability p and chooses $a \in \{x,y\}$, her utility is given by

$$U(a,p;r) = \begin{cases} u(p)+r & \text{if} \quad a=x \\ u(1-p) & \text{if} \quad a=y. \end{cases} \tag{3.5}$$

If choosing x, she receives a state-independent *remuneration* $r \geq 0$[10] and belief utility $u(p)$ for believing that her choice x is harmless for the other agent with probability p. The belief utility u is weakly increasing, weakly concave, and continuously differentiable; we normalize $u(1)=0$. That is, the dictator feels no disutility if he is certain that the action of his choice does not harm the receiver.[11] If choosing y, she only receives belief utility $u(1-p)$ for believing that her choice y is harmless for the other with probability $1-p$. We call u the *(other-regarding) preference type* of the agent.

Before deciding between x and y, the agent has unrestricted access to information about the state at no cost.[12] Formally, she can choose any *signal structure*, i.e., a joint distribution of a set of signals $s \in S$ and the state. For any signal structure, the distribution of her posterior beliefs $\Pr(X|s)$ conditional on the realized signal s must be Bayes-plausible.[13] In the following, we model her choice of a signal structure as the choice of a posterior belief distribution $\tau \in \Delta(\Delta(\Omega))$ from the set of Bayes-plausible distributions.[14] After choosing τ, a posterior $p \in \text{supp}(\tau)$ is drawn by nature and privately observed by the agent. Then, she decides on $a \in \{x,y\}$ to maximize her utility given the realized posterior belief p.

Preliminaries. A posterior belief determines the utility in two steps. First, it determines the agent's choice of action between x and y – the agent chooses the action

9. The negativity of the externality is only a matter of normalization. Our model applies to all situations where one of the options is better for the other agent, and one worse.

10. A remuneration $r > 0$ might arise in situations where she receives a choice-contingent monetary payment, e.g. a commission, a prize or it might arise from choice-contingent non-monetary rewards, e.g. an increase in the reputation within a group or the feeling of satisfaction from a particular choice.

11. This normalization is without loss of generality. Our results hold as long as u is weakly increasing, weakly concave, and continuously differentiable.

12. Later, in the Online Supplement, we describe how the model naturally generalizes to the situation when information is costly.

13. A distribution of posteriors is called Bayes-plausible if the expected posterior equals the prior, that is $\sum_{p \in supp(\tau)} p \Pr_\tau(p) = p_0$.

14. It has been shown in the literature on Bayesian persuasion (Kamenica and Gentzkow, 2011) that the model where the agent can choose any signal structure and the model where she can choose any Bayes-plausible distribution of posteriors are equivalent. It is because, for any Bayes-plausible distribution of beliefs τ, there is a signal structure such that the distribution of posterior belief $\Pr(X|s)$ is τ.

that maximizes $U(a, p; r)$ for any given belief. Then, together with the chosen option, it determines the utility. Hence, we can get rid of the argument a in the utility function and directly express utility as a function of posterior belief p:

$$V(p; r) = \max_{a \in \{x,y\}} U(a, p; r). \tag{3.6}$$

$V(p; r)$ is the continuation value for any posterior realization p given remuneration r. The optimal posterior belief distribution τ is the one maximizing her expected continuation value, i.e.

$$E_\tau V(p; r) = \sum_{p \in \text{supp}(\tau)} \Pr_\tau(p) V(p; r). \tag{3.7}$$

Before analyzing the optimal posterior belief distribution, we first narrow down the space of the optimal τ in Lemma 1. It shows that for any other-regarding preference type, there exists an optimal posterior distribution τ^* that is supported on *two* (potentially identical) beliefs $\underline{p} \leq \bar{p} \in [0, 1]$. We show in the proof of Lemma 1 that the agent chooses x when her posterior belief realizes as $\bar{p}$ and chooses y when her posterior realizes as $\underline{p}$, whenever $\underline{p} < \bar{p}$. The proof is in the Appendix.

Lemma 1. For any $r \geq 0$ and any u, there is an optimal posterior distribution τ^* with binary support $\text{supp}(\tau^*) = \{\underline{p}, \bar{p}\}$ and $\underline{p} \leq p_0 \leq \bar{p} \in [0, 1]$, where p_0 is the agent's prior belief.

3.3.2 The Optimal Information Acquisition Strategy

In this section, we compare the optimal belief cutoffs $\underline{p}$ and $\bar{p}$ in the scenario without a remuneration ($r = 0$) and in the scenario with a remuneration ($r > 0$). Theorem 1 shows that, for *all* preference types that acquire at least some information in both scenarios, both belief cutoffs $\underline{p}$ and $\bar{p}$ are weakly smaller when there is a remuneration for option x, i.e. $r > 0$.

Theorem 1. Take any u and any optimal belief cutoffs $(\underline{p}^{co}, \bar{p}^{co})$ given $r = 0$. If it is not optimal to acquire no information given $\bar{r} > 0$, then for any optimal belief cutoffs $(\underline{p}^{tr}, \bar{p}^{tr})$ given $\bar{r} > 0$,

$$\underline{p}^{tr} \leq \underline{p}^{co}, \tag{3.8}$$

$$\bar{p}^{tr} \leq \bar{p}^{co}. \tag{3.9}$$

The formal proof is in the Appendix, while we elaborate the intuition below. Note that the statement of the theorem is trivially true when the agent's belief utility u is weakly convex: no matter $r = 0$ or $r > 0$, she always strictly prefers accurate beliefs and acquires all possible information, i.e. $\underline{p}^{tr} = \underline{p}^{c} = 0$ and $\bar{p}^{tr} = \bar{p}^{c} = 1$. Our empirical finding that most of the dictators stop acquiring information when their beliefs are far from certainty suggests that the belief utility is likely concave. The

concavity of the belief utility captures the following psychological mechanism: it is increasingly more uncomfortable for the individual to choose an option, as she becomes more certain that her chosen option is the one worse for the other.

The intuition of this theorem goes back to the observation that only an agent in scenario $r > 0$ wants to persuade herself to choose x. Lower $\overline{p}$ or lower $\underline{p}$ makes the agent to choose x with a higher probability. Recall that she chooses x only if the upper cutoff $\overline{p}$ is realized. To increase the probability that she chooses x, she has to increase the probability that the upper cutoff $\overline{p}$ is realized. There are two things that our Bayesian agent can do to increase the probability of the upper cutoff being realized. First, she can require lower certainty to choose x, i.e., reduce the upper cutoff $\overline{p}$, such that it is realized with higher probability. Second, she can require higher certainty to choose y, i.e., reduce the lower cutoff $\overline{p}$, such that the lower cutoff is realized with lower probability.

While the model does not restrict the information environment of the agent, our experiment focuses on information environments in which the decision-makers can sequentially acquire noisy information and freely decide when to stop. The feature of unrestricted access to information in our model approximates these information environments. The optimal belief cutoffs $(\underline{p}, \overline{p})$ of our model translate into the following dynamic behaviour in the experiment: a dictator chooses *belief cutoffs* $\underline{p}$ and $\overline{p}$ and acquires information until her belief reaches either $\overline{p}$ or $\underline{p}$. She then chooses x if $\overline{p}$ is reached, and y if $\underline{p}$ is reached.[15] Theorem 1 is in line with our empirical findings from the laboratory experiment, i.e. when most of the information received so far indicates that the remunerative option x is *harmless* to others ($p > p_0$), weakly more dictators in *Tradeoff* ($r > 0$) stop acquiring information; when most of the information received so far indicates that the remunerative option x is *harmful* to others ($p < p_0$), weakly less dictators in *Tradeoff* ($r > 0$) stop acquiring information.

In finer details, Theorem 1 can be understood by considering the optimal information acquisition strategy in scenario $r = 0$ and $r > 0$ respectively. We first introduce an important value of belief

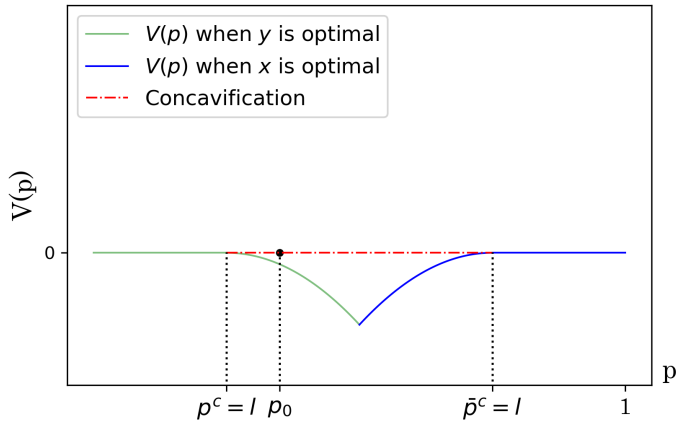
$$l = \min \{q : u(q) = 0\}.$$

$l \leq 1$ is the threshold above which any further certainty that her chosen option is harmless does not increase her belief utility any more;[16] whereas when her belief that her chosen option is harmless is lower than l, her utility increases when she gains additional certainty that the option is harmless. We term l the agent's *moral*

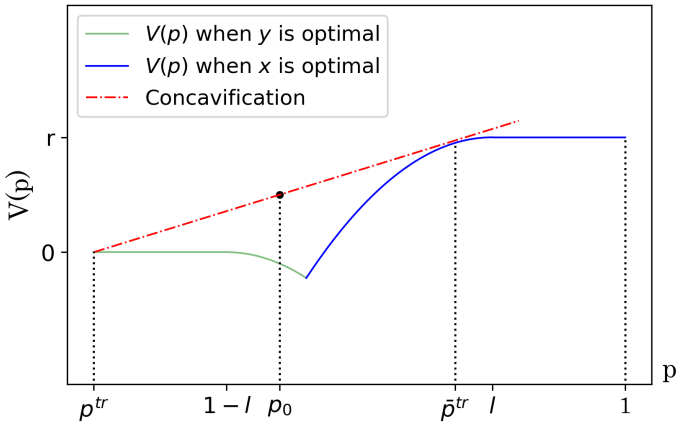
15. Note that an equivalence between static persuasion models and dynamic information acquisition models in the presence of information cost has been shown formally in Morris and Strack (2019).

16. See Simon (1955) for seminal literature on satisficing. One feature of the satisficing behavior in our setting is that the agent exhibits satisficing behavior for beliefs instead of outcomes.

standard. A moral standard $l < 1$ captures the idea of satisficing – the agent is "satisficed" if she is certain with l probability that her chosen option does not harm others, and any further certainty no longer brings her additional utility.



(a) $l < 0$ and $r = 0$



(b) $l < 0$ and $r > 0$

Figure 3.5. Illustration of Optimal Cutoffs

Figure 3.6a illustrates $V(p)$ in the scenario without remuneration, i.e. $r = 0$. The agent's only concern is her belief utility u. Whenever she is more certain than her moral standard that her decision does not harm the other ($\underline{p}^c < 1 - l$ or $\overline{p}^c > l$), her belief utility is at its highest value 0. Therefore, any information acquisition strategy that always makes her more certain in the state than her moral standard is optimal for her. Formally:

Theorem 2. When $r = 0$, any cutoff pair $(\underline{p}^c, \overline{p}^c)$ with $\underline{p}^c \in [0, 1-l]$ and $\overline{p}^c \in [l, 1]$ is optimal.

Next, we turn to the optimal cutoffs when x is remunerative, i.e., $r > 0$.

In the presence of remuneration, the agent values her belief utility that she does no harm on the other, as well as the remuneration utility r given by choosing x. Regarding $\underline{p}^{tr}$, we first observe that the agent chooses the non-remunerative option y only if she is certain that y is the option harmless to the other, i.e. $\underline{p}^{tr} = 0$. This is because if choosing y for any $\underline{p}^{tr} > 0$, she can always improve her belief utility by choosing y at $\underline{p}^{tr} = 0$. Meanwhile, $\underline{p}^{tr} = 0$ minimizes the probability that $\underline{p}^{tr}$ is realized, *ceteris paribus*, so that she can choose the remunerative option x with the highest probability. Theorem 2 shows that $\underline{p}^c \in [0, 1-l]$. Therefore, like Theorem 1 shows, $\underline{p}^{tr} \leq \underline{p}^c$.

Regarding $\overline{p}^{tr}$, when she considers choosing the remunerative option x, she faces a trade-off: on the upside, she appreciates higher $\overline{p}^{tr}$, as it increases her belief utility from believing that x is harmless with higher certainty; on the downside, the higher $\overline{p}^{tr}$ is, the lower is the probability that it is realized, and hence the lower is the probability that she can choose x. This tradeoff determines the optimal cutoff $\overline{p}^{tr}$.

Among those who acquire information, there are two classes of agent types. The types in the first class acquire complete information given $r > 0$.[17] These types must have moral standard $l = 1$, i.e., they are not satisfied by any belief lesser than certainty. Besides, they value additional certainty in their beliefs so much that their marginal belief utility still exceeds the remuneration r even when their belief is already very close to certainty. Since their moral standard is equal to 1, these types also acquire complete information when $r = 0$. Hence for them, $\overline{p}^{tr} = \overline{p}^c = 1$, consistent with Theorem 1.

The second class of types does not acquire complete information. Next, we will show that for the types who do not acquire complete information, $\overline{p}^{tr} < \overline{p}^c$, in line with Theorem 1. We first formally express the tradeoff between the belief utility and the risk of the undesirable realization of $\underline{p}^{tr}$. Recall the agent's maximization problem (3.7); given that $\underline{p}^{tr} = 0$[18] and $V(0) = 0$, her maximization problem is

$$\overline{p}^{tr} = argmax_{p \in [p_0, 1]} \Pr(p) V(p; r), \tag{3.10}$$

subject to the Bayes-plausibility constraint. Bayes-plausibility, together with $\underline{p}^{tr} = 0$, implies that $\Pr(\overline{p}^{tr}) = \frac{p_0}{\overline{p}^{tr}} \in [0, 1]$. The first-order condition is therefore

$$\begin{aligned} & Pr(p)u'(p) + \frac{\partial Pr(p)}{\partial p} V(p; r) = 0 \qquad (3.11) \\ \Leftrightarrow \quad & \frac{p_0}{p} u'(p) - \frac{p_0}{p^2} V(p; r) = 0 \end{aligned}$$

17. Formally they are those who satisfy the condition $u'(1) \geq r$. $u'(1) \geq r$ implies that $l = 1$. Since u is continuously differentiable, if $l < 1$, then $r > u'(1) = 0$.

18. Proof see Appendix.

$\bar{p}^{tr}$ is the solution to (3.11). The intuition of (3.11) is that its first term describes the marginal increase in belief utility u for being more certain that the chosen option x is harmless; and its second term captures the marginal undesirable risk that higher information can make the remunerative option unacceptable.

Figure 3.6b illustrates the problem geometrically. It is easy to see in (3.11) that its solution $\bar{p}^{tr}$ is where the linear line connecting point $(0,0)$ and point $(\bar{p}^{tr}, V(\bar{p}^{tr}))$ is exactly tangential to $V(p)$ at the latter. This linear line is the smallest concave function lying weakly above $V(p)$, which we refer to as the concavification of $V(p)$. The expected utility given belief cutoffs $(0,\bar{p}^{tr})$ is given by the intersection of the concavification and the vertical line above p_0.

Theorem 3 shows that when the interior solution $\tilde{p}$ exists, it must be the optimal upper cutoff $\bar{p}^{tr}$ and it must be smaller than l, i.e. $\bar{p}^{tr} = \tilde{p} < l$.[19] Since $\bar{p}^{co} \geq l$, for this class of types, $\bar{p}^{tr} < \bar{p}^{co}$.

Theorem 3. When $r > 0$, for any type u with $u'(1) < r$, let $\tilde{p}$ be the interior solution of (3.11). When $p_0 \geq \tilde{p}$, the agent acquires no information; When $p_0 < \tilde{p}$, $\underline{p}^{tr} = 0$ and $\bar{p}^{tr} = \tilde{p} < l$.

The proof of Theorem 3 is in the Appendix.

In summary, we have shown that, just like Theorem 1 states, (i) the lower cutoff $\underline{p}^{tr}$ is always weakly smaller than $\underline{p}^{co}$, since it is always 0; (ii) the upper cutoff $\bar{p}^{tr}$ is always weakly smaller than $\bar{p}^{c}$. Specifically, when the upper cutoff $\bar{p}^{tr}$ is equal to 1, $\bar{p}^{co}$ also must be 1; when the upper cutoff $\bar{p}^{tr}$ is smaller than 1, it must be strictly smaller than $\bar{p}^{co}$.

3.3.3 Belief Cutoffs in the Experiment

In this section, we infer the belief cutoffs using the experimental data and compare them between treatments.

We find that the large majority of subjects behave consistently with the model (431 out of 496; *Control*: 228 out of 246; *Tradeoff*: 203 out of 250), i.e., they choose x if they stop at a posterior weakly above the prior or y if they stop at a posterior weakly below the prior.[20] For dictators who stop acquiring information at the equivalent of their prior belief of 0.35, 0.35 is interpreted as their upper cutoffs if they choose x and as their lower cutoffs if they choose y.

Table 3.7 summarizes the fraction of dictators who stop at their upper belief cutoffs $\bar{p}$. It reveals that the distribution of posterior belief cutoffs differ between *Tradeoff* and *Control*. In *Tradeoff*, 49% dictators stop acquiring information at a

19. In the Appendix we show that an interior solution of (3.11) exists when $u'(1) < r$, i.e. whenever the agent does not acquire full information.

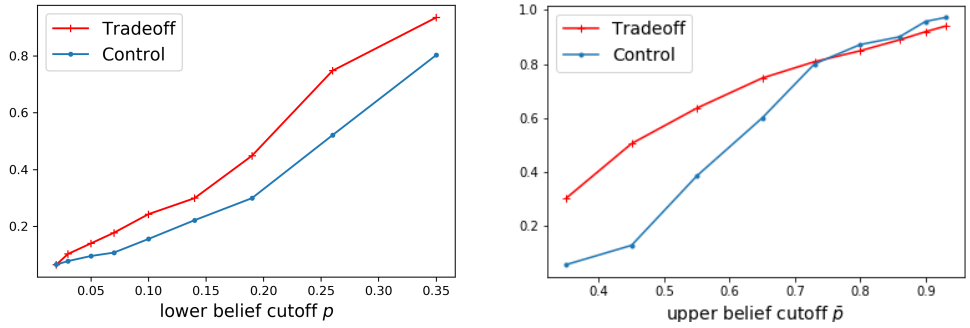
20. In the *Control* treatment, 14 dictators choose y after having received more good news, 4 dictator choose x after having received more bad news. In the *Tradeoff* treatment, 10 dictators choose y after having received more good news, 37 subjects choose x after having received more bad news.

posterior belief above the prior belief, while in *Control* the fraction is only 31% (Chi square, $p = 0.00$). This finding is consistent with our theoretical model.

Figure 3.6 shows the empirical cumulative distribution function of the upper and lower belief cutoff. Both the lower and the upper cutoff are *lower* in the Tradeoff treatment, consistent with Theorem 1.[21]

Table 3.7. Proportion of Dictators Reaching the Upper Belief Cutoff $\bar{p}$

	Overall	*Tradeoff*	*Control*	Chi-2 p
Stop at $\bar{p}$	39%	49%	31%	0.00



(a) CDf of the lower belief cutoff

(b) CDF of the upper belief cutoff

These figures show the empirical cumulative distribution functions of the lower belief cutoff (Figure 3.7a) and the upper belief cutoff (Figure 3.7b). The CDF of the lower belief cutoff reflects the data of dictators who stop information acquisition at posterior beliefs weakly below the prior and choose y. The CDF of the upper belief cutoff reflects dictators who stop weakly above the prior and choose x.

Figure 3.6. Distribution of the Observed Belief Cutoffs

3.4 Receiver Welfare

Do the dictators in *Tradeoff*, for whom x is self-rewarding, more often choose the option that reduces the receivers' payment, than the dictators in *Control*? This might seem to be the case, since the dictators in treatment *Tradeoff* might bias towards choosing x, whereas the dictators in *Control* are impartial between the option x and y. Indeed, our data show that in both states the dictators in *Tradeoff* are more likely to choose x than dictators in *Control* (details see Section 3.4.2). However, we find

21. For the interpretation of the right tail of the distribution, note that only 5% dictators stop at beliefs higher than 0.80, such that only very few observations drive the estimation of the cumulative distribution functions at very high beliefs.

that, despite the higher tendency to choose x, the dictators in *Tradeoff* do *not* choose the option that reduces the receivers' payments significantly more often (*Tradeoff*: 32%; *Control*: 27%; Chi-2 $p = 0.17$). These two observations seem to contradict each other. What is the missing piece of the puzzle?

An option being remunerative does not only directly affect the agent's decision between the options (*the decision effect*), but also indirectly by affecting how she acquires information (*the information effect*). In Section 3.4.1, we theoretically show that, while the decision effect of the remuneration is always negative on the welfare of the other, the information effect is positive for some agent types. This information effect can sometimes offset the decision effect and leads to an overall neutral or even positive effect of the remuneration on the welfare of the other.

This counter-intuitive result arises from a moral hazard problem: when impartial between options, the agent might acquire little information. Therefore she sometimes mistakenly chooses the harmful option because she is ill-informed about the state. We show that one option being remunerative can mitigate this moral hazard problem. Although she now more often falsely chooses x, the agent *less* often falsely chooses y because she now requires higher certainty in the innocuousness of y before choosing it.

In Section 3.4.2, we take the theory to our data. We disentangle the decision and the information effect in our experimental data. In our experiment, the information effect indeed improves receiver welfare, and it offsets the negative decision effect, resulting in no overall significant difference between treatments regarding the proportion of receivers whose payments were reduced by the dictators' decisions.

3.4.1 Disentangling the Decision Effect and the Information Effect in Theory

In this section, we theoretically analyze the effect of a remuneration $\bar{r} > 0$ for option x on the welfare of the passive person affected by the agent's decision between x and y. We call the passive person "the other".

First, let us formally express the expected utility of the other. Let $v(a, \omega)$ be the utility of the other when the agent chooses $a \in \{x, y\}$ in $\omega \in \{X, Y\}$. Recall that the other has negative utility of -1 if the chosen option does not match the state and has utility 0 otherwise, i.e. $v(x, Y) = v(y, X) = -1$ and $v(x, X) = v(y, Y) = 0$.

For any given belief, the agent chooses the option $a \in \{x, y\}$ that maximizes her own utility $U(a, p; r)$ (see (3.5)). Hence for given r, we write the chosen option as a function of her belief p, i.e. $a_r(p) = \max_{a \in [x,y]} U(a, p; r)$. We call a_r the *decision rule* given r. τ pins down the joint distribution of the posterior belief realization and the state ω, given the prior belief p_0 and the Bayes-plausibility constrain. We hence can write the expected utility of the other given posterior belief distribution τ as

$$E_\tau v \equiv E v(a_r(p), \omega) | \tau. \tag{3.12}$$

Notice in (3.12) that the agent determines the expected utility of the other by making two decisions: first, she chooses the decision rule $a_r(p)$; second, she chooses the information acquisition strategy and hence the posterior belief distribution τ. A remuneration for choosing x, i.e. $\bar{r} > 0$, affects both decisions. We call the effect of $\bar{r}$ on $E_\tau v$ through changing the decision rule $a_r(p)$ the *decision effect*; and we call the one through changing the posterior belief distribution τ the *information effect*.

We write the overall effect of the remuneration $\bar{r}$ on the expected utility of the other as

$$Ev(a_{tr}(p), \omega)|\tau^{tr} - Ev(a_{co}(p), \omega)|\tau^{co}, \tag{3.13}$$

where τ^{tr} is the optimal information acquisition strategy given $\bar{r} > 0$ and τ^{co} the optimal strategy given $r = 0$; a^{tr} is the decision rule when $\bar{r} > 0$ and a^{co} the decision rule given $r = 0$.

Next, we discuss the decision effect and the information effect of $\bar{r}$ on the expected utility $Ev(a_r(p), \omega)|\tau$ respectively.

The Decision Effect. We first discuss the decision effect, i.e. the effect of $\bar{r} > 0$ on the expected utility of the other through affecting the agent's choice between x and y given posterior beliefs. This effect can be expressed by the difference of expected utility of the other when keeping the posterior belief distribution fixed at τ^{co} and changing the decision rule:

$$\text{DE} \equiv Ev(\boldsymbol{a_{tr}}(p), \omega)|\tau^{co} - Ev(\boldsymbol{a_c}(p), \omega)|\tau^{co} \tag{3.14}$$

For any belief p, the agent chooses x over y iff

$$u(p) + r > u(1 - p). \tag{3.15}$$

When there is no remuneration, i.e. $r = 0$, for any belief the agent always chooses the option that is less likely to be harmless for the other. She is indifferent between the two options at belief $p = 0.5$. However, when x is remunerative, i.e., $\bar{r} > 0$, the agent's indifferent point becomes lower – she chooses x for less certainty that it is harmless. Theorem 4 shows that this change of decision rule makes the other weakly worse off. The proof is in the Appendix.

Theorem 4. For any $\bar{r} > 0$, any agent type $u(.)$,

$$Ev(\boldsymbol{a_{tr}}(p), \omega)|\tau^{co} \leq Ev(\boldsymbol{a_{co}}(p), \omega)|\tau^{co},$$

i.e. the decision effect is weakly negative.

The Information Effect. Next we discuss the effect of remuneration $\bar{r} > 0$ on the expected utility of the other that is due to change of the agent's optimal information acquisition strategy τ, i.e. the information effect. This effect can be expressed by

keeping fixed the decision rule that is optimal given $\bar{r} > 0$ and changing the information acquisition strategy from τ^{co} to τ^{tr}:

$$\text{IE} \equiv Ev(a_{tr}(p), \omega)|\boldsymbol{\tau^{tr}} - Ev(a_{tr}(p), \omega)|\boldsymbol{\tau^{co}}. \tag{3.16}$$

Recall that Theorem 2 shows that when $r = 0$, an agent with moral standard $l < 1$ has optimal information acquisition strategy that does not yield perfect beliefs, i.e., cutoffs other than 0 and 1 can be optimal for the agent. It implies that when $r = 0$, if the agent is satisfied before her belief reaches certainty, a *true* moral hazard problem arises: the agent only acquires partial information about the state. Consequently, she sometimes mistakenly chooses x when the state is Y, and she also sometimes mistakenly chooses y when the state is X.

Theorem 5 shows that a self-reward of an option can serve as a motivation device and mitigate the moral hazard problem. The intuition is that when x is remunerative, the agent makes no mistakes when she chooses option y – she only chooses y when she is certain that it is the option harmless to the other. The information effect, therefore, can be positive. We also show that the positive information effect can dominate the decision effect and result in an overall positive effect of $\bar{r} > 0$ on the expected utility of the other.

Theorem 5. There are agent types u such that the presence of a remuneration $\bar{r} > 0$ has a positive information effect on the expected utility of the other and the overall effect of $\bar{r} > 0$ on the expected utility of the other is positive.

The proof is in the Appendix. Note that the overall effect is

$$\text{DE} + \text{IE} = Ev(a_{tr}(p), \omega)|\boldsymbol{\tau^{tr}} - Ev(a_{co}(p), \omega)|\boldsymbol{\tau^{co}},$$

namely, the difference of the other's expected utility between $\bar{r} > 0$ and $r = 0$.

3.4.2 The Receiver Welfare in the Experiment

In this section, we discuss the receiver welfare in our laboratory experiment. A direct between-treatment comparison of the receiver welfare confounds two effects of the self-reward on the receiver welfare: first, it directly affects their decision between x and y, given any acquired information (decision effect); second, the self-reward affects dictators' information acquisition, which in turn affects their beliefs about the unknown state and their choices between the options (information effect). Before we disentangle the decision effect and the information effect, we first present the dictators' choice of x and y given realized posterior beliefs at their decisions.

We observe in our data that, fixing the posterior belief, the dictators in *Tradeoff* decide differently between x and y than the dictators in *Control*. This difference

affects the proportion of receivers whose incomes are reduced by the dictators' decisions (the decision effect). Specifically, in both treatments most dictators who have received more good news than bad news choose option x (*Tradeoff*: 91%, *Control*: 93%, chi-2 $p = 0.63$). Difference arises among those who have received more bad news than good news – a significantly higher fraction of these dictators in *Tradeoff* choose option x (*Tradeoff*: 27%, *Control*: 3%, chi-2 $p = 0.00$). Similarly, among those who have received equal number of good and bad news (final belief on x being harmless $= 0.35$), including those who acquire no information, significantly more dictators in treatment *Tradeoff* choose x than those in the *Control* treatment (*Tradeoff*: 81%, *Control*: 11%, chi-2 $p = 0.00$).

To empirically disentangle the decision effect and the information effect of the remuneration on the receivers' welfare, we construct a *Counterfactual* scenario, in which dictators acquire information as in the *Control* treatment, but decide as in the *Tradeoff* treatment given the acquired information and the final posterior beliefs (as illustrated in Table 3.8). When comparing the receiver welfare in the *Counterfactual* to the *Control* treatment, we isolate the decision effect by keeping fixed the information acquisition behavior; when comparing the receiver welfare in the *Counterfactual* to that in the *Tradeoff* treatment, we isolate the information effect by keeping fixed the decision between x and y given beliefs.

Table 3.8. *Counterfactual* Scenario

	Control	*Tradeoff*
posterior beliefs	×	
decision given belief		×
compared to the *Counterfactual*	*decision effect*	*information effect*

Tables 3.9a and 3.9b show the decision effect and the information effect respectively. In Table 3.9a, we compare the *Counterfactual* with the *Control* and find a negative decision effect. The dictators in the *Counterfactual*, who employ the decision rules in *Tradeoff* given any posterior belief, choose x more often in both states. Overall, in the *Counterfactual*, the proportion of unharmed receivers is lower than in the *Control* treatment (62% compared to 73%). This means that the decision effect is negative: option x being self-rewarding for the dictators leads to a change of decision rule that makes the receivers worse-off.

In Table 3.9b, we compare *Tradeoff* with the *Counterfactual* and find a positive information effect. The remuneration makes a higher fraction of dictators choose x when x is harmless (81% compared to 75%), and a higher fraction of dictators to choose y when y is harmless (60% compared to 54%). Overall, in *Tradeoff*, the proportion of unharmed receivers is higher than in the *Counterfactual* (68% compared to 62%). The information effect of remuneration on the receiver welfare is

hence positive: option x being self-rewarding makes the dictators acquire information strategically, and in turn, improves the receiver welfare.

As discussed before, there is a moral hazard problem when no option is remunerative – the dictators do not fully learn the state before they make a decision and hence often mistakenly choose the harmful option for the receiver. Note that in both states, the proportions of dictators who choose the harmless option for the receiver are lower in *Counterfactual* than in *Tradeoff*. This difference can only be due to different information acquisition behavior since the decision rule is the same between the *Counterfactual* and *Tradeoff*. In our experiment, in *Control*, 36% dictators who are actually in the *Good* state stop acquiring information at a belief in the Good state lower than their prior. The additional payment that the dictators can obtain by choosing x mitigates this moral hazard problem: in treatment *Tradeoff*, the proportion of dictators in the Good state who stop acquiring information below the prior is 26% – lower than in *Control*.

Aggregating both effects, the proportion of the receivers spared from harm does not significantly differ between the *Tradeoff* and the *Control* (68% compared to 73%, Chi-2 $p = 0.17$). It is decreased from 73% (*Control*) to 62% (*Counterfactual*) by more selfish decision-making, i.e. the decision effect, and is increased from 62% (*Counterfactual*) to 68% (*Tradeoff*) by strategic information acquisition, i.e. the information effect.

Table 3.9. The Effects of Remuneration on Receiver Welfare

(a) The Decision Effect

State	Good State (x harmless)	Bad State (y harmless)	Overall
Counterfactual:			
% no harm	75%	54%	62%
(# total dictators)	(88)	(158)	(246)
Control:			
% no harm	54%	83%	73%
(# total dictators)	(88)	(158)	(246)
The decision effect:			-11%

(b) The Information Effect

State	Good state (x harmless)	Bad state (y harmless)	Overall
Tradeoff:			
% no harm	81%	60%	68%
(# total dictators)	(87)	(163)	(250)
Counterfactual:			
% no harm	75%	54%	62%
(# total dictators)	(88)	(158)	(246)
The information effect:			6%

This table presents the decision effect and the information effect of the remuneration in our experiment. The *Counterfactual* is calculated by combining the posterior beliefs from the *Control* and the mapping from beliefs to choices in the dictator game from *Tradeoff*. Comparing the *Conterfactual* to the *Control* (*Tradeoff*), we obtain the decision effect (information effect).

3.5 Concluding Remarks

This paper experimentally and theoretically investigates how people acquire information about the externalities of their options before making a decision.

We present experimental evidence that when faced with a self-benefiting option that might harm others, individuals acquire noisy information strategically: they tend to carry on acquiring information when they have received mostly information suggesting that the selfish decision harms others; while they tend to stop having received information indicating the opposite. Moreover, in our experiment, individuals with higher intelligence exhibit a stronger tendency to acquire information this way, suggesting that this information acquisition behavior is more likely to be due to strategic considerations than limited cognitive ability.

This empirical finding sheds light on how people acquire information in various contexts where decisions incur unknown consequences on others, and noisy information is available for inquiry. One example is the credence goods market. The research on credence goods has been focusing on the deceptive behavior of the credence goods provider, while the psychology of them is less understood. The credence goods providers – physicians, car mechanics, taxi drivers – who care for the well-being of their customers face a dilemma between their monetary compensations and their unwillingness to harm their customers. Our finding suggests that credence goods providers might mitigate this dilemma by strategically learning about the best option for the customers. If by examining the need of a customer, they can persuade themselves that a profitable option is the right one for the customer, their dilemma is resolved.

Our findings also help to understand labor market discrimination. A discriminatory recruiter who likes to think of himself as nondiscriminatory might be able to maintain his positive self-view while hiring in a biased manner, by selectively stopping interviewing the candidate to persuade himself that a candidate of his less preferred character is disqualified. This insight has implications on the quality distribution of successful labor market candidates across ethnic groups and gender.

In other contexts, such as charitable giving and media consumption for voting, our result highlights the importance of the first pieces of information sent and received. If the potential donors' first information about a charitable organization is negative, she might readily stop learning about the charity and decide to keep her money in her pocket. The charity will then have a hard time to raise for its cause. When a voter inquires about an ethical issue with a personal cost for him (e.g., additional taxes), if the first news articles that he reads lean against it, the voter is likely to stop the inquiry and vote against it.

In terms of theory, we propose a tractable model that analyzes the acquisition of information with any degree of noise. The model applies techniques developed for studying interpersonal Bayesian persuasion, by Kamenica and Gentzkow (2011), to the investigation of information acquisition of a single Bayesian agent. It offers

intuitive geometric tools that generate rich results. Our model also addresses an important yet little understood dimension of social decision making: other-regarding preferences under uncertainty. We suggest that other-regarding preferences can be modeled by a belief utility that is increasing in the probability with which the agent believes that her decision does not harm others. With this modeling approach, we can explain many empirical findings on the information choices in social decisions with uncertainty, including the noisy information acquisition strategy found in our experiment and the avoidance of perfect information observed by Dana, Weber, and Kuang (2007) and Feiler (2014) (see Appendix 3.A.6). We hope that it is a step towards a more comprehensive understanding of other-regarding preferences, and might facilitate modeling in related settings in the future.

Our finding that motivated information acquisition can improve the welfare of the other affected by the decision is particularly relevant for policymakers. Under the opposite intuition that strategic information acquisition motivated by selfish incentives must increase negative externalities, it might seem to be a good idea to de-bias the information acquisition behavior by involving an independent investigator whose compensation is not related to the decision. However, our model and our data suggest that sometimes such strategic information acquisition motivated by selfish incentives can make the other party affected by the decision *better-off*. This finding offers the novel insight that assigning the job of collecting information to an independent investigator, who is disinterested in the decision, can sometimes lead to worse decision making and more negative externalities.

3.A Empirical Appendices

3.A.1 Summarizing Statistics

Here we provide summarizing statistics on our data. The basic information of the subjects in each treatment is summarized in Table 3.10.

Table 3.10. Basic Information of Subjects

		no. obs.	Good State	women	student	av. age
Force	Tradeoff	82	.34	.45	.95	22
	Control	79	.37	.54	.95	22
	p value		.73	.24	.56	.50
NoForce	Tradeoff	168	.35	.66	.93	24
	Control	167	.35	.65	.92	24
	p value		.97	.79	.56	.36
Pooled	Tradeoff	250	.35	.59	.94	24
	Control	246	.36	.61	.93	24
	p value		.82	.62	.56	.25

This table presents the basic characteristics of our subjects in each treatment. The Mann-Whitney U test verifies that our randomization was successful.

3.A.2 Number of Balls Drawn and the Posterior Beliefs

Table 3.11 summarizes the dictators' information acquisition behavior.

Table 3.11. Information Acquisition Behavior

		no. balls (median)	av. belief at decision	% stop above prior
Force	Tradeoff	7.5	.30	.33
	Control	4	.36	.37
	p value	.04	.04	.67
NoForce	Tradeoff	5	.34	.37
	Control	6	.33	.33
	p value	.92	.76	.44
Pooled	Tradeoff	6	.35	.36
	Control	5	.36	.34
	p value	.24	.82	.71

This table presents the statistics of the dictators' information acquisition behavior and the Mann-Whitney-U test p values comparing between *Tradeoff* and *Control*, respectively. In the *NoForce* treatments, only dictators who draw at least one ball are included.

3.A.3 Dictator Game Decision

Table 3.12 summarizes the dictator game decisions.

Table 3.12. Dictator Game Decisions

		Choosing x%			Harm %
		Good	Bad	Overall	
Force	Tradeoff	.71	.43	.54	.38
	Control	.62	.14	.32	.23
	p value	.46	.00	.01	.04
NoForce	Tradeoff	.86	.38	.55	.30
	Control	.51	.18	.29	.29
	p value	.00	.00	.00	.84
Pooled	Tradeoff	.81	.40	.54	.32
	Control	.54	.16	.30	.27
	p value	.00	.00	.00	.17

The first three columns of this table presents the proportions of dictators who choose x given *Good* and *Bad* states and the treatments, together with the Mann-Whitney U test p values comparing between *Tradeoff* and *Control* respectively. In the *Good* state, x does not harm the receiver, while in the *Bad* state it does.The last column presents the percentage of dictators whose decision reduced the receivers' payoffs in the dictator game.

3.A.4 Robustness Check: The Logistic Regression

Using the data at the person-draw level, we estimate the following logistic model as a robustness check and find result similar to that in Section: 3.2.2.2.

$$logit\, h(X) = X \cdot b + Z \cdot a + (C + T \cdot c), \tag{3.17}$$

where $h(X)$ is the probability that the dictator stops acquiring information after that draw; X denotes the same covariates of interest as in the Cox model, i.e.

$$X \cdot b = \beta_1 \text{Tradeoff} + \beta_2 \text{Info} + \beta_{12} \text{Tradeoff} \times \text{Info}.$$

The control valuables in Z include gender, cognitive ability, prosociality and belief accuracy, all measured in the same way as in the Cox model in Section 3.2.2.2. T is a vector of time dummies, which captures the time dependency of the probability to stop acquiring information.

When interpreting the results, this logistic model can be viewed as a hazard model in which the covariates proportionally affect the *odds* of stopping acquiring information (Cox, 1975). Formally,

$$\begin{aligned} \frac{h(t)}{1-h(t)} &= \frac{h_0(t)}{1-h_0(t)} \cdot exp(X_t \cdot b + Z_t \cdot a) \\ \Rightarrow \underbrace{log\left(\frac{h(t)}{1-h(t)}\right)}_{logit\, h(X)} &= \underbrace{log\left(\frac{h_0(t)}{1-h_0(t)}\right)}_{C+T\cdot c} + X_t \cdot b + Z_t \cdot a. \end{aligned} \tag{3.18}$$

Unlike in the framework of the Cox model, the coefficients here cannot be interpreted as hazard ratios. Instead, they should be interpreted as odds ratios. Our prediction that the hazard to stop acquiring information is lower in *Tradeoff* when bad news dominates suggests a negative β_1. And the prediction that the hazard is higher when good news dominates suggests a positive $\beta_1 + \beta_{12,Good}$. Results reported in Table 3.13 support these predictions.

Table 3.13. The Logistic Model Results

		Pooling All		Force	NoForce
$\hat{\boldsymbol{\beta}}_{\mathbf{1}}$	treatment *Tradeoff*	-.25*	-.26*	-.56**	-.18
		(.15)	(.15)	(.25)	(.18)
$\hat{\boldsymbol{\beta}}_{\mathbf{12}}$	Tradeoff ×				
	Good news dominance	.35*	.37*	.71**	.34
		(.22)	(.22)	(.37)	(.26)
	Balanced	-.54	-.53	-.62	-.40
		(.40)	(.41)	(.73)	(.49)
$\hat{\beta}_2$	Good news dominance	-.21	-.21	-.14	-.26
		(.18)	(.18)	(.29)	(.24)
	Balanced	-.67**	-.68**	-.46	-78**
		(.28)	(.28)	(.46)	(.35)
Control Variables:					
Belief Accuracy		Yes	Yes	Yes	Yes
Gender, IQ, Prosociality		Yes	Yes	Yes	Yes
Time Dummies		Yes	Yes	Yes	Yes
Force Treatment Dummy		No	Yes	–	–
Observations (person-draws)		4,658	4,658	1,567	2,932
Pseudo R2		.07	.07	.09	.07

This table presents the estimated *coefficients* of the logistic model, with standard errors clustered at the individual level. *, **, and *** denote significance at the 10, 5, and 1 percent level. The dependent variable is the hazard to stop acquiring information, and the key coefficients of interests are $\hat{\beta}_1$ and $\hat{\beta}_{12}$. $\exp(\hat{\beta}_1)$ reflects the treatment effect on the dictators' odds to stop acquiring further information, given information histories dominated by bad news. And $\exp(\hat{\beta}_1 + \hat{\beta}_{12}|\text{Good news dominance})$ reflects the treatment effect on the odds, given information histories dominated by good news. We control for belief accuracy, gender, the prosocial types (categorized by the SVO test), and the cognitive ability (measured by in Raven's matrices test). The time dependency of the odds is accounted for by including a dummy for each period.

3.A.5 Complementary Stage

After the experiment, we elicited the dictators' posterior beliefs on the state and their SVO scores. We also asked them to answer a questionnaire consisting of questions on their sociodemographics, self-reported risk preferences, time preferences, preferences for fairness, reciprocity. A selective subset of the HEXACO personality inventory (Ashton and Lee, 2009) and five items from Raven's progressive matrices intelligence test are also included.

Elicited Beliefs. In the experiment, we display the Bayesian posterior belief on the state after each draw of information on the screens of the dictators. After the dictators stop acquiring information, we elicit subjects' beliefs on the state, given all the information acquired. Figure 3.7 plots the histogram of the difference between the Bayesian posterior beliefs, and the elicited posterior beliefs at the end of the information acquisition. The majority of subjects' elicited beliefs coincide with the Bayesian posterior beliefs after the last ball they draw (299 out of 496), the elicited beliefs of the self-rewarding option x being harmless are higher than the Bayesian posterior beliefs by 2.60% (one-sample t-test $p = 0.00$). Figure 3.7 reveals no systematic bias in the elicited beliefs.

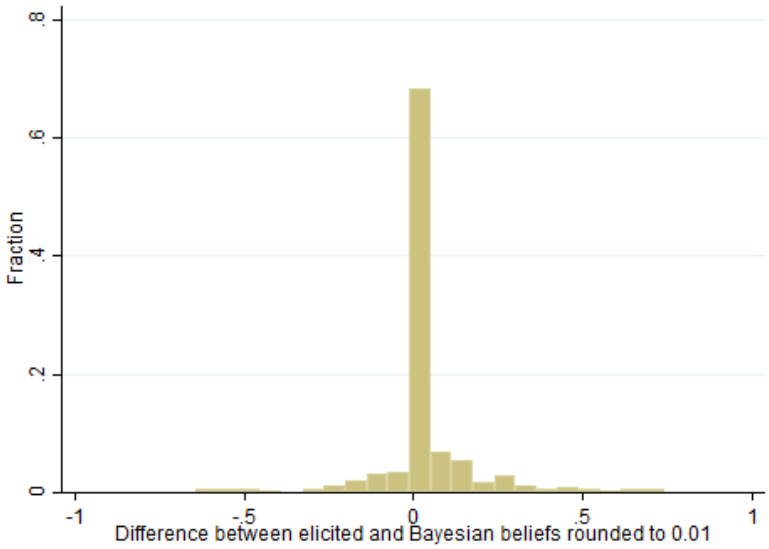


Figure 3.7. Difference between elicited posterior beliefs and Bayesian posterior beliefs

SVO Scores. The average SVO score of all the subjects is 20.49, with no significant difference between *Tradeoff* and *Control* treatments (Mann-Whitney-U test, $p = 0.84$). According to Murphy, Ackermann, and Handgraaf (2011), 48% subjects are categorized as "prosocials", 15% "individualists" and 37% "competitive type".

Cognitive Abilities. On average, the subjects answered 3.60 out of 5 questions in Raven's matrices test correctly. There is no significant difference between *Control* and *Tradeoff* treatments (Chi-square $p = 0.12$). When asked about a simple question on probability, in both treatments 92% subjects answer correctly (Mann-Whitney-U test $p = 0.85$).[22]

Preferences. To elicit risk preferences, time preferences, preferences for fairness, and reciprocity, we use survey questions in Falk, Becker, Dohmen, Huffman, and Sunde (2016). We report the exact questions in Table 3.14. All answers are given on a 0 to 10 scale.

HEXACO-60 proposed by Ashton and Lee (2009) is a personality inventory that assesses the following six personality dimensions: Honesty-Humility (HH), Emotionality (EM), Extraversion (EX), Agreeableness (AG), Conscientiousness (CO), and Openness to Experiences (OP). We select 4 questions with the highest factor loading in each dimension (as reported in Moshagen, Hilbig, and Zettler, 2014) and in addition, include 4 questions from the Altruism versus Antagonism scale (AA) proposed in Lee and Ashton (2006). Table 3.15 reports the exact questions we ask. All questions are answered on a scale from 1 to 5, where 5 means strongly agree, and 1 means strongly disagree. We use the German self-report form provided by hexaco.org.

22. We use the following question to elicit subjects' understanding of probabilities:
Imagine the following 4 bags with 100 fruits in each. One fruit will be randomly taken out. For which bag, the probability of taking a banana is 40%?
A. A bag with 20 bananas.
B. A bag with 40 bananas.
C. A bag with 0 banana.
D. A bag with 100 bananas.
The correct answer is B.

Table 3.14. Preferences Elicitation in the Questionnaire

Preferences for	Question
Risk	Please tell me, in general, how willing or unwilling you are to take risks. (10 means very willing, 0 means completely unwilling)
Time	How willing are you to give up something beneficial for your today to benefit more from that in the future? (10 means very willing, 0 means completely unwilling)
Altruism	I am always ready to help others, without expecting anything in return.
Fairness	Q1: I think it is very important to be fair. Q2: I, in general, agree that unfair behaviors should be punished.
Positive reciprocity	I am always ready to go out of my way to return a favor.
Negative reciprocity	I am always ready to take revenge if I have been treated unfairly.

Table 3.15. Selected Items From the HEXACO Personality Inventory

Dimension	Question
HH	12. If I knew that I could never get caught, I would be willing to steal a million dollars. 18. Having a lot of money is not especially important to me. 42. I would get a lot of pleasure from owning expensive luxury goods. 60. I'd be tempted to use counterfeit money if I were sure I could get away with it.
EM	17. When I suffer from a painful experience, I need someone to make me feel comfortable. 41. I can handle difficult situations without needing emotional support from anyone else. 47. I feel strong emotions when someone close to me is going away for a long time. 59. I remain unemotional even in situations where most people get very sentimental
EX	10. I rarely express my opinions in group meetings. 22. On most days, I feel cheerful and optimistic. 28. I feel that I am an unpopular person. 40. The first thing that I always do in a new place is to make friends.
AG	3. I rarely hold a grudge, even against people who have badly wronged me. 15. People sometimes tell me that I'm too stubborn. 21. People think of me as someone who has a quick temper. 45. Most people tend to get angry more quickly than I do.
CO	2. I plan and organize things, to avoid scrambling at the last minute. 26. When working, I sometimes have difficulties due to being disorganized. 44. I make a lot of mistakes because I don't think before I act. 56. I prefer to do whatever comes to mind, rather than stick to a plan.
OP	1. I would be quite bored by a visit to an art gallery. 13. I would enjoy creating a work of art, such as a novel, a song, or a painting. 25. If I had the opportunity, I would like to attend a classical music concert. 55. I find it boring to discuss philosophy.
AA	97. I have sympathy for people who are less fortunate than I am. 98. I try to give generously to those in need. 99. It wouldn't bother me to harm someone I didn't like. 100. People see me as a hard-hearted person.

3.A.6 Additional Theoretical Results

In this section, we discuss two additional results of our model and the respective empirical evidence: the avoidance of noisy and perfect information. While our experiment focuses on *noisy* information, the information that can be analyzed in our model encompasses both noisy and *perfect* information, i.e., information that reveals the truth in one piece.

Our model predicts that with or without a remunerative option, there are agents who acquire no noisy information at all (Section 3.A.6.1). In line with this predic-

tion, in both treatments in our experiment, some dictators do not acquire any noisy information before making the dictator decision.

Regarding perfect information, our model predicts that there are agents who avoid *perfectly revealing* information. This result is consistent with the empirical finding of Dana, Weber, and Kuang (2007). Besides, we theoretically show that the higher is the prior belief that the self-rewarding option is harmless to others, the more agent types would avoid perfect information. This prediction is in line with the experimental finding of Feiler (2014).

3.A.6.1 Avoidance of Noisy Information

Our model predicts that both in decisions with or without a remunerative option, some agent types move on to the decision without acquiring any noisy information (Theorem 6).

Theorem 6. 1. When $r = 0$, for any prior $p_0 \in (0,1)$, there is a set $S^{co}(p_0)$ of preference types u that avoid information completely, i.e. the belief cutoffs $\underline{p}^{co} = \overline{p}^{co} = p_0$ are optimal.

2. When $r > 0$, for any prior $p_0 \in (0,1)$, there is a set $S^{tr}(p_0)$ of preference types u that avoid information completely, i.e. the belief cutoffs $\underline{p}^{tr} = \overline{p}^{tr} = p_0$ are optimal.

In the experiment, we find that 15% and 7% dictators do not acquire any noisy information in the *Tradeoff–NoForce* and the *NoForce –Control* treatment respectively (Chi-2 $p = 0.00$). In the *Tradeoff–NoForce* treatment, among those who avoid noisy information completely 96% choose the remunerative action x (25/26). In contrast, in the *Control–Force* treatment, only 17% of those who avoid noisy information choose x (2/12).

In theory, the types of the agent who acquire no information, when no option is remunerative, are those with moral standard $l \leq p_0$ or $l \leq 1 - p_0$, i.e., those for whom there is already no gain in belief utility for more certain beliefs at the prior belief. Recall that we fix the dictators' prior belief in our experiment at 35% in x's innocuousness. The observation that in the *Control* treatment, most dictators who avoid noisy information completely choose option y suggests that these are the dictators with moral standards $l \leq 65\%$. They are satisfied with 65% certainty that y is the harmless option, and more certainty does not bring them any additional utility.

In the decision with remuneration, the agent decides not to acquire noisy information only if she would choose x at the prior belief. The further information then poses an undesirable risk that it might reverse her decision from x to y. She avoids noisy information only when this risk outweighs her utility gain from more certain beliefs that she does not harm the other. This intuition is consistent with

the observation that all dictators who avoid noisy information completely in the *Tradeoff–NoForce* treatment choose the option x, except for one.

3.A.6.2 Avoidance of Perfect Information

While our experimental investigation focuses on the acquisition of noisy information that unravels the unknown state piece by piece, our model also makes predictions about how people acquire information that reveals the truth at once – *perfectly revealing information*.

Recall that in our theoretical model, the agent can choose any signal structure (Section 3.3.1). Perfectly revealing information is a special case of the signal structures that the model encompasses. Let $p_0 \in (0,1)$ be any uncertain prior belief. The decision whether or not to acquire a piece of perfectly revealing information is formally the preference between the posterior belief distribution τ^{p_0} that has mass 1 on the prior belief p_0 and the posterior distribution τ^{ce} with $supp(\tau^{ce}) = \{0,1\}$. Theorem 7 shows that in the presence of remuneration, for any uncertain prior belief, there are types of dictators who would avoid perfectly revealing information. The higher is the prior belief in the alignment between the dictator's and the receiver's payment, the more types of dictators would avoid perfect information.

Theorem 7. 1. When $r > 0$, for any prior $p_0 \in (0,1)$, there is a set $S(p_0)$ of preference types u that avoid perfectly revealing information, i.e. $\tau^{p_0} \succ \tau^{ce}$.

2. For any prior beliefs $p_0^l < p_0^h \in (0,1)$, it holds that $S(p_0^l) \subset S(p_0^h)$.

A piece of perfectly revealing information either makes the agent certain that the remunerative option is harmless, or makes her certain that it is harmful. For an agent who would choose the remunerative option at the prior belief, if the realized signal is that the remunerative option is harmless, the agent gains in belief utility, as she becomes more certain that she is not harming the other. But on the other hand, she faces the risk that the realized signal would make her certain that the remunerative option is harmful so that she would have to forgo the remuneration and choose the other option instead.

The first item of Theorem 7 shows that for any uncertain prior belief, there are some types of agents for whom the risk of having to forgo the remuneration outweighs the potential gain in belief utility so that they would rather avoid the perfect information and make a decision based on their prior beliefs. Trivially, agent types with weakly convex preference type u will always acquire perfect information. These agents who avoid perfect information must have strictly concave belief utility u.

The second item of Theorem 7 predicts that when the prior belief is higher, it is optimal for more agent types to avoid the perfect information and choose the remunerative option directly. When the prior belief increases, on the one hand, the additional belief utility from being certain that the preferred option is indeed harm-

less decreases, so the perfect information becomes less attractive; but on the other hand, the probability that the remunerative option is harmless increases, so the perfect information becomes more attractive. Since these agent types who avoid perfect information have strictly concave belief utilities u, the magnitude of the first negative effect becomes larger with increasing prior, while the magnitude of the second positive effect is linear in the prior belief. Therefore, as the prior increases, the perfect information becomes overall less attractive and more agent types would avoid perfect information.

These predictions are consistent with previous empirical findings. In a dictator environment similar to ours, Dana, Weber, and Kuang (2007) find that a significant fraction of dictators avoids information that reveals the *ex-ante* unknown state all at once. Feiler (2014) further documents that the fraction of dictators who avoid such perfectly revealing information increases with the dictators' prior belief that a self-benefiting option has no negative externality.

3.A.7 Proofs

3.A.7.1 Proof of Lemma 1 and Theorem 2

Proof of Lemma 1. The statement holds trivially when u is strictly convex since then the agent strictly prefers Black-well more informative information and the unique optional posterior distribution has support on $\underline{p} = 0$ and $\overline{p} = 1$. It remains to prove the lemma when u is weakly concave. Consider any optimal posterior distribution τ. Suppose that there are two beliefs $p_1, p_2 \geq p_0$ with $\Pr_\tau(p_1) > 0, \Pr_\tau(p_2) > 0$. Let $\hat{p} = \Pr_\tau(p_1) + \Pr_\tau(p_2)p_2$. Then $\hat{p} \geq p_0$. Also,

$$V(\hat{p}) - (\Pr_\tau(p_1)V(p_1) + \Pr_\tau(p_2)V(p_2)) = u(\hat{p}) - (\Pr_\tau(p_1)u(p_1) + \Pr_\tau(p_2)u(p_2)) \geq 0,$$

since u is weakly concave. So, we see that she is weakly better off with the posterior distribution that arises from τ when shifting the mass from p_1 and p_2 to $\hat{p}$. Suppose that there are two beliefs $p_1, p_2 \leq p_0$ with $\Pr_\tau(p_1) > 0, \Pr_\tau(p_2) > 0$. The analogous argument shows that shifting mass from p_1 and p_2 to $\hat{p} = \Pr_\tau(p_1) + \Pr_\tau(p_2)p_2$ makes her weakly better off. This finishes the proof of the Lemma.

Proof of Theorem 2. When $r = 0$, any pair of beliefs $(\underline{p}^c, \overline{p}^c)$ with $\underline{p}^c \in [1-l, l]^c$ and $\overline{p}^c \in [1-l, l]^c$ implies an expected continuation value $\mathrm{E}_{(\underline{p}^c, \overline{p}^c)}V(p)$ of 0. Since, given $r = 0$, the expected continuation value for any posterior distribution τ is weakly negative, any such pair of belief cutoffs is optimal. This finishes the proof of Theorem 2.

3.A.7.2 Proof of Theorem 6, Theorem 1, and Theorem 3

Let $r > 0$. Any optimal pair of belief cutoffs $\underline{p} \leq \overline{p}$ satisfies Bayes-plausibility,

$$\overline{p}\Pr_\tau(\overline{p}) + \underline{p}(1 - \Pr_\tau(\underline{p})) = p_0, \tag{3.19}$$

which pins down how likely it is that she stops at the upper cutoff $\overline{p}$ and how likely it is that she stops at the lower cutoff $\underline{p}$, given the prior belief. The likelihood of the upper belief cutoff is negatively proportional to its relative distance to the prior,

$$\Pr_\tau(\overline{p}) = \frac{p_0 - \underline{p}}{\overline{p} - \underline{p}}. \tag{3.20}$$

The expected continuation value, given belief cutoffs $(\underline{p}, \overline{p})$ is therefore

$$\mathrm{E}_{(\underline{p},\overline{p})} V(p) = \frac{p_0 - \underline{p}}{\overline{p} - \underline{p}} V(\overline{p}) + \frac{\overline{p} - p_0}{\overline{p} - \underline{p}} V(\underline{p}), \tag{3.21}$$

which is simply the value of the affine function connecting $V(\overline{p})$ and $V(\underline{p})$ through the prior. Since $r > 0$, there is a unique pair of beliefs $(\underset{\sim}{p}, \tilde{p})$ that support the concave envelope.[23] Note that

$$\underset{\sim}{p} = 0. \tag{3.22}$$

The following lemma shows that the pair of belief cutoffs $(\underset{\sim}{p}, \tilde{p})$ is the unique optimal strategy whenever it is not optimal to acquire no information.

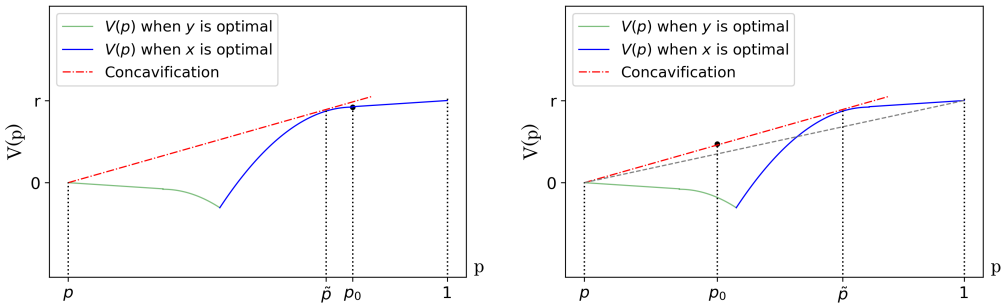
Lemma 2. Let $r > 0$.

1. When $p_0 \in [\underset{\sim}{p}, \tilde{p})$, then there is a unique pair of optimal belief cutoffs, given by $(\underline{p}^{tr}, \overline{p}^{tr}) = (\underset{\sim}{p}, \tilde{p})$.
2. When $p_0 \notin [\underset{\sim}{p}, \tilde{p})$, then acquiring no information is optimal, i.e. the belief cutoffs $\underline{p}^{tr} = \overline{p}^{tr} = p_0$ are optimal.

Proof. Consider any two belief cutoffs $\underline{p} \leq p_0 \leq \overline{p}$ and the value of the connecting function at the prior. The optimal belief cutoffs maximize (3.21).
The claim can be seen geometrically: when $p_0 \in [\underset{\sim}{p}, \tilde{p}]$, the optimal belief cutoffs are given by the unique pairs of beliefs $\underline{p}$ and $\tilde{p}$ that support the concave envelope of V (see Figure 3.9a). Whenever $p_0 \notin [\underset{\sim}{p}, \tilde{p})$, the maximum of (3.21) is achieved through no information acquisition (see Figure 3.9b). □

Proof of Theorem 6. The first item of Theorem 6 follows from Theorem 2 since, for any prior $p_0 \in (0,1)$, there is an open set of preference types u for which $p_0 \in [l, 1-l]^c$.
Lemma 2 together with (3.22) implies that for $r > 0$, the optimal lower belief cutoff is $\underset{\sim}{p} = 0$. Therefore, it follows from Lemma 2 that, for any prior $p_0 \in (0,1)$, the set $S(p_0)$ of types u for which no information acquisition is optimal is given by the types for which $p_0 \geq \tilde{p}$. This shows the second item of Theorem 6. Also note that this set is strictly smaller when the prior is larger.

23. The smallest concave function that lies weakly above V is called the concave envelope of V; compare to Figure 3.8.



(a) Complete Information Avoidance **(b)** Optimal Belief Cutoffs Other Than the Prior

Note: The green and the blue line show the continuation value function V which is defined component-wise. The beliefs $\underset{\sim}{p}$ and $\tilde{p}$ are the unique beliefs supporting the concave envelope of V. The agent cannot improve on the belief cutoffs $(\underset{\sim}{p}, \tilde{p})$ when $p_0 \in [\underset{\sim}{p}, \tilde{p}]$: This can be seen geometrically: given (3.21), the optimal cutoffs maximize the value of the connecting function at the prior. Any other choice implies that the connecting line (grey) takes a value at the prior lower than $V(p_0)$ (left) or lower than the line connecting $\underset{\sim}{p}$ and $\tilde{p}$ (right). When $p_0 \notin [\underset{\sim}{p}, \tilde{p}]$, it is optimal for the agent to acquire no information.

Figure 3.8. Optimal Belief Cutoffs

Proof of Theorem 1. It remains to show the theorem for the case when u is weakly concave; see the discussion after Theorem 2. Take any weakly concave u, any optimal strategy $(\underline{p}^c, \overline{p}^c)$ given $r = 0$. Suppose that it is not optimal to acquire no information given $\bar{r} < 0$.

Given Lemma 2, there are unique optimal belief cutoffs $\underline{p}^{tr} < p_0 < \overline{p}^{tr}$, and, given (3.22), it holds $\underline{p}^{tr} = 0$. Given (3.21), the upper belief cutoff $\overline{p}^{tr}$ maximizes

$$\max_{p \in [p_0, 1]} \Pr(p) V(p; \bar{r}) \tag{3.23}$$

subject to the Bayes-plausibility constraint that $\Pr(p)p = p_0$. Plugging in the Bayes-plausibility constraint gives the objective function

$$\max_{p \in [p_0, 1]} \frac{p}{p_0} V(p; \bar{r}), \tag{3.24}$$

and taking derivatives gives the first-order condition

$$\begin{aligned} &\frac{p_0}{p} u'(p) - \frac{p_0}{p^2} V(p; \bar{r}) = 0 \\ \Leftrightarrow \quad &p u'(p) - V(p; \bar{r}) = 0. \end{aligned} \tag{3.25}$$

The maximization problem (3.23) has a solution since continuous functions take maxima on compact sets. Note that the second derivative of the objective function (3.24) is weakly negative,

$$\frac{\partial}{\partial p}(p u'(p) - u(p) - \bar{r}) = p u''(p) \leq 0, \tag{3.26}$$

where we used that u is weakly concave.

Case 1. $u'(1) \geq \bar{r} > 0$

The condition $u'(1) \geq \bar{r}$ implies that $l = 1$. Therefore, without remuneration, $r = 0$, the optimal belief cutoffs are $(\underline{p}^c, \overline{p}^c) = (0, 1)$. Since $\underline{p}^{tr} = 0$, the inequalities (3.8) and (3.9) follow. This finishes the proof of the theorem for Case 1.

Case 2. $u'(1) < \bar{r}$

The condition $u'(1) < \bar{r}$ is equivalent to

$$1u'(1) - V(1) < 0. \tag{3.27}$$

If $l < 1$, then $u'(p) = 0$ for $p \geq l$ since u is continuously differentiable. In any case,

$$lu'(l) - V(l) < 0. \tag{3.28}$$

Suppose that the derivative of the objective function is weakly negative for all $p \in [p_0, 1]$; this is equivalent to

$$p_0 u'(p_0) - V(p_0; \bar{r}) \leq 0, \tag{3.29}$$

given (3.26). Then, the objective function is maximized at the boundary $\overline{p} = p_0$. Bayes-plausibility implies that $\Pr(\overline{p}) = 1$. We conclude, that no information acquisition is optimal. However, we excluded this case by assumption. Therefore,

$$p_0 u'(p_0) - V(p_0; \bar{r}) > 0, \tag{3.30}$$

and it follows from the intermediate value theorem, (3.28), and (3.30) that the first-order condition (3.25) is satisfied by some $\tilde{p}$ with $p_0 < \tilde{p} < l$. It follows from (3.26) that the derivative of the objective function is weakly positive for $p < \tilde{p}$ and weakly negative for $p > \tilde{p}$ such that $\tilde{p}$ maximizes the objective function. We conclude that the belief cutoffs

$$(\underline{p}^{tr}, \overline{p}^{tr}) = (0, \tilde{p}) \tag{3.31}$$

are optimal. Moreover, given (3.28), any optimal upper belief cutoff satisfies $\overline{p}^{tr} < l$ and the first-order condition.

Now, we finish the proof of the theorem for Case 2. The inequality (3.9) follows directly from $\underline{p} = 0$. Theorem 2 states that, without remuneration, the optimal belief cutoffs are the pairs of beliefs $(\underline{p}^c, \overline{p}^c)$ that satisfy $\overline{p}^c \geq l$ and $\underline{p} \leq 1 - l$. Since $\tilde{p} < l$, the inequality (3.8) holds strictly. When $l < 1$, then, without remuneration, there are optimal belief cutoffs $(\underline{p}^{co}, \overline{p}^c)$ with $\underline{p}^c > 0 = \underline{p}^{tr}$. Hence, the inequality (3.8) holds strictly. This finishes the proof of the theorem.

Proof of Theorem 3. See the proof of Theorem 1 above.

3.A.7.3 Proof of Theorem 4 and Theorem 5

Proof of Theorem 4. Let p^* solve

$$u(p) + \bar{r} = u(1-p). \tag{3.32}$$

It is easy to see that $p^* < 0.5$. Recall that $\overline{p}^c \geq l$ and $\underline{p}^c \leq 1-l < 0.5$. If $1-l < p^*$, then at the posterior belief cutoffs $\overline{p}^c$ and $\underline{p}^c$, the agent chooses x and y according to decision rule a_{tr} just like according to a_c. There is no decision effect and $Ev(\boldsymbol{a_{tr}}(p),\omega)|\tau^{co} = Ev(\boldsymbol{a_c}(p),\omega)|\tau^{co}$.
If $1-l \geq p^*$, then for any lower belief cutoff $\underline{p}^c \in [0,p^*]$, there is also no the decision effect because the agent chooses x at $\overline{p}^c$ and y at $\underline{p}^c$.
However, for any lower belief cutoff $\underline{p}^c \in [p^*, 1-l]$, the agent chooses x at $\underline{p}^c$ instead if they decide according to a_{tr}. Therefore with a_{tr}, the expected utility of the other if the lower cutoff is realized is

$$-Pr(Y|\underline{p}^c) = -(1-\underline{p}^c). \tag{3.33}$$

Whereas with a_c, the expected utility of the other if the lower cutoff is realized is

$$-Pr(X|\underline{p}^c) = -\underline{p}^c. \tag{3.34}$$

Recall $\underline{p}^c < 0.5$, hence $-(1-\underline{p}^c) < -\underline{p}^c$, i,e, the expected utility of the other if the lower cutoff is realized is lower with a_{tr} than with a_c. Since the probability that the lower cutoff is realized is pinned down by τ^c, and the expected utility of the other if the upper cutoff is realized is the same between the scenario with a_c and the scenario with a_{tr}, the expected utility of the other is strictly *lower* keeping τ^c fixed and changing a_c to a_{tr}. The decision effect is strictly negative, i.e. $Ev(\boldsymbol{a_{tr}}(p),\omega)|\tau^{co} < Ev(\boldsymbol{a_c}(p),\omega)|\tau^{co}$.

Proof of Theorem 5. We prove the theorem with an example. Consider u such that $p_0 < l < 1$ and $u''(x) \to \infty$ for $p \in [1-l, l]$. When $r = 0$, it follows from Theorem 2 that a pair of optimal cutoffs are $\underline{p}^{co} = 1-l$ and $\overline{p}^{co} = l$. When $r > 0$, $\underline{p}^{tr} = 0$ and $\overline{p}^{tr} \to l$, since these two points support the concave envelope.

First, we prove that for this agent, the information effect is strictly positive.

What does the agent choose at each cutoff? It follows from (3.15) that, for any $r > 0$, the belief p^* where she is indifferent between x and y, i.e. $u(p^*) + r = u(1-p^*)$, converges to $\frac{1}{2}$. So the agent chooses y at the two lower cutoffs $\underline{p}^{tr}$ and $\underline{p}^{co}$, and x at the two upper cutoffs $\overline{p}^{tr}$ and $\overline{p}^{co}$. That is, the decision effect (3.14) converges to zero.

Let us now derive the expected utility of the other with τ^{co} and τ^{tr}, when fixing $a_{tr}(p)$.

First, with τ^{co},

$$\begin{aligned} Ev(a_{tr}(p),\omega)|\tau^{co} &= -1 \cdot Pr(\underline{p}^{co})Pr(X|\underline{p}^{co}) + (-1) \cdot Pr(\overline{p}^{co})Pr(Y|\overline{p}^{co}) && (3.35)\\ &= -(Pr(\underline{p}^{co})(1-l) + Pr(\overline{p}^{co})(1-l)) && (3.36)\\ &= -(Pr(\underline{p}^{co}) + Pr(\overline{p}^{co}))(1-l) && (3.37)\\ &= -(1-l). && (3.38) \end{aligned}$$

Then with τ^{tr},

$$\begin{aligned} Ev(a_{tr}(p),\omega)|\tau^{tr} &\to -1 \cdot Pr(\underline{p}^{tr}Pr(X|\underline{p}^{tr}) + (-1) \cdot Pr(\overline{p}^{tr})Pr(Y|\overline{p}^{tr}) && (3.39)\\ &= -(Pr(\underline{p}^{tr}) \cdot 0 + \frac{p_0}{l}(1-l)) && (3.40)\\ &= -\frac{p_0}{l}(1-l). && (3.41) \end{aligned}$$

Since $p_0 < l$,

$$Ev(a_{tr}(p),\omega)|\tau^{tr} > Ev(a_{tr}(p),\omega)|\tau^{co}, \quad (3.42)$$

In other words, we have proven that for this agent if $p_0 < l$, the information effect (3.16) is strictly positive.

We have already discussed above that the decision effect converges to zero for this agent. We hence can conclude that the overall effect is strictly positive. It finishes the proof.

3.A.7.4 Proof of Theorem 7

Proof of Theorem 7. Item 1 of Theorem 7 is a corollary of the second item of Theorem 6: if a preference type prefers no information over all possible information structures, clearly, she prefers no information over the fully revealing signals.
We prove item 2 of Theorem 7 by contradiction.
For any prior belief $p_0^l < p_0^h \in (0,1)$, if an agent type prefers the prior to the fully revealing signals at this prior then

$$\begin{aligned} \tau^{p_0^l} &\succ \tau^{ce} && (3.43)\\ \Leftrightarrow r + u(p_0^l) &> rp_0^l. && (3.44) \end{aligned}$$

Suppose this agent prefers the fully revealing signals to prior p_0^h, then

$$\begin{aligned} \tau^{p_0^h} &\prec \tau^{ce} && (3.45)\\ \Leftrightarrow r + u(p_0^h) &< rp_0^h. && (3.46) \end{aligned}$$

Subtract 3.44 from 3.46 and rearrange,we get:

$$\frac{u(p_0^h) - u(p_0^l)}{p_0^h - p_0^l} < r. \quad (3.47)$$

Since $u(\cdot)$ is concave and $p_0^l < p_0^h < 1$,

$$\frac{u(1) - u(p_0^h)}{1 - p_0^h} < r. \tag{3.48}$$

Since $u(1) = 0$, we get

$$\begin{aligned} \frac{-u(p_0^h)}{1 - p_0^h} &< r, && (3.49)\\ \Rightarrow r + u(p_0^h) &> r p_0^h, && (3.50)\\ \Rightarrow \tau^{p_0^h} &\succ \tau^{ce}. && (3.51) \end{aligned}$$

Contradiction. Hence $\tau^{p_0^h} \succeq \tau^{ce}$. In other words, $S(p_0^l) \subset S(p_0^h)$.

3.A.8 An Order of Other-Regarding Preferences

This section shows how the preference model of Section 3.3.1 allows for stable comparative predictions about differences in the information acquisition behaviour of two decision-makers.
Theorem 1 and Theorem 5 illustrate that decision-makers acquire information about the consequences of their choices on others in a self-deceptive way. Some types exhibit a very strong form of self-deception, they avoid information completely; see Theorem 5. Others exploit information in the following sense: the optimal belief cutoffs $(\underline{p}^{tr}, \overline{p}^{tr})$ with remuneration $\bar{r} > 0$ are weakly smaller than any optimal belief cutoffs $(\underline{p}^{c}, \overline{p}^{c})$ without remuneration, i.e. $r = 0$; see Theorem 1. Among those, some types exploit the information less strongly than others, meaning that the differences in the optimal belief cutoffs with and without remuneration, i.e. $\underline{p}^c - \underline{p}^{tr}$ and $\overline{p}^c - \overline{p}^{tr}$ are smaller.

The next result shows that there is a simple ordering other-regarding preference types that translates into an ordering of the predicted degree of self-deceptive behaviour: the lower the curvature of the belief utility, the more self-deceptive the agent behaves across *all* possible situations. We say that a preference type with belief utility u is *more self-deceptive* than a type with belief utility v if $u' < v'$ and write $u \succ_{dec} v$. For any type u, let

$$\underline{\delta}(u; \bar{r}) = \max \left[\underline{p}^{co}(u) - \underline{p}^{tr}(u) \right], \tag{3.52}$$

$$\overline{\delta}(u; \bar{r}) = \max \left[\overline{p}^{co}(u) - \overline{p}^{tr}(u) \right]. \tag{3.53}$$

where we take the maximum over all pairs of optimal belief cutoffs $(\underline{p}^{co}(u), \overline{p}^{co}(u))$ given $r = 0$ and all pairs pf optimal belief cutoffs $(\underline{p}^{tr}(u), \overline{p}^{tr}(u))$ given $\bar{r}$.

Theorem 8. Let $u \succ_{dec} v$. Then for all $\bar{r} > 0$, the following holds.

1. If it is optimal for the v-type to avoid information completely given $\bar{r}$, then, this is also true for the u-type. The converse is not true.
2. If it is not optimal for the v-type to avoid information completely given $\bar{r}$, then either it is optimal for the u-type to avoid information completely given $\bar{r}$ or

$$\overline{\delta}(u;\bar{r}) > \overline{\delta}(v;\bar{r}),$$
$$\underline{\delta}(u;\bar{r}) \geq \underline{\delta}(v;\bar{r}).$$

Note that, given the normalization $u(1) = v(1) = 0$, the relation $u \succ_{dec} v$ implies that $v(0) < u(0)$. It follows from (3.5) that under certainty about the state $\omega = B$, type v chooses the other-regarding action y whenever u does. We see that the ordering $\succ_{dec}$ is an extension of the natural ordering of other-regarding preference types under certainty.

Proof. Let $u \succ_{dec} v$ and consider the situation with remuneration $\bar{r} > 0$. Let $\underset{\sim}{p}(u)$ and $\tilde{p}(u)$ be the unique pair of beliefs supporting the concave envelope of the continuation value function V of the u-type and $\underset{\sim}{p}(v)$ and $\tilde{p}(v)$ be the unique pair of beliefs supporting the concave envelope of the continuation value function V of the v-type. Recall that $\underset{\sim}{p}(u) = \underset{\sim}{p}(v) = 0$, given Lemma 2 and (3.22).
Consider the first item of the theorem. Lemma 2 says that it is optimal for the v-type to avoid information completely, given $\bar{r}$, if and only if $p_0 \notin [\underset{\sim}{p}(v), \tilde{p}(v)]$. Similarly, it is optimal for the u-type to avoid information completely, given $\bar{r}$, if and only if $p_0 \notin [\underset{\sim}{p}(u), \tilde{p}(u)]$. Since $\underset{\sim}{p}(u) = \underset{\sim}{p}(v) = 0$, to prove the first item of the theorem it suffices to show that

$$\tilde{p}(u) < \tilde{p}(u). \tag{3.54}$$

Note that the beliefs $\underset{\sim}{p}(u)$ and $\tilde{p}(u)$ supporting the concave envelope of V satisfy

$$V(\tilde{p}(u);\bar{r}) - V(\underset{\sim}{p}(u);\bar{r}) = u'(\tilde{p})(\tilde{p}(u) - \underset{\sim}{p}(u). \tag{3.55}$$

Since $\underset{\sim}{p}(u) = \underset{\sim}{p}(v) = 0$ and $V(0,\bar{r}) = 0$, this implies that $\tilde{p}(u)$ satisfies the condition

$$pu'(p) - V(p;\bar{r}) = 0; \tag{3.56}$$

compare to the first-order condition (3.25). Similarly, $\tilde{p}(v)$ satisfies

$$pv'(p) - V(p;\bar{r}) = 0; \tag{3.57}$$

Therefore, if

$$\begin{aligned} \forall p : pu'(p) - V(p;\bar{r}) &> pv'(p) - V(p;\bar{r}), \\ \Leftrightarrow \forall p : pu'(p) - u(p) &> pv'(p) - v(p), \end{aligned} \tag{3.58}$$

then (3.54) holds. We rewrite (3.58) using $u(1) = v(1) = 0$,

$$\forall p : pu'(p) + \int_p^1 u'(p)dp > pv'(p) + \int_p^1 v'(p)dp. \tag{3.59}$$

Clearly, $u' > v'$ implies (3.59). This finishes the proof of the first item.

Consider the second item of the theorem. Given Lemma 2, if it is not optimal for the type to avoid information completely, then, $p_0 < \tilde{p}(v)$ and the optimal belief cutoffs of the v-type given $\bar{r}$ are $\underset{\sim}{p}(v) = 0$ and $\tilde{p}(v)$. Given (3.54), we have to distinguish two cases.

Case 1. $\tilde{p}(u) \leq p_0$

Then, it follows from given Lemma 2 that it is optimal for the u-type not to acquire any information. This finishes the proof of the second item in this case.

Case 2. $p_0 < \tilde{p}(u) < \tilde{p}(v)$

Then, it follows from Lemma 2 that the optimal belief cutoffs of the u-type given $\bar{r}$ are $\underset{\sim}{p}(u) = 0$ and $\tilde{p}(u)$. Consider $l(v) = \min\{p \in [0,1] : v(p) = 0\}$ and $l(u) = \min\{p \in [0,1] : u(p) = 0\}$. Note that $u \succ_{dec} v$ implies $l(u) > l(v)$. The claim of the second item of the theorem follows from the characterization of the optimal belief cutoffs without remuneration in Theorem 1 and from (3.54). □

3.A.9 Parametric Examples

Isoelastic Belief Utility. Let $u(p) = -\alpha(1-p)^n$ for some $n > 1$. The parameter α moderates how an individual with perfect knowledge would value the interest of others relative to her own. The parameter n captures how the agent values the interest of others when his belief changes; formally, n is the elasticity of the belief utility function as a function of $q = 1 - p$, that is as a function of the belief that the option is harmful to the other.[24]

Given Theorem 1, the agent's optimal belief cutoffs are weakly smaller with remuneration $\bar{r} > 0$ than without if it is not optimal to avoid information completely given $\bar{r}$. Note that $u'(1) = 0$ for all $n > 1$ such that it follows from the proof of Theorem 1 (see Case 2, in particular (3.30)) that it is not optimal to avoid information completely if

$$\begin{aligned} 0 &< p_0u'(p_0) - u(p_0) - r \\ \Leftrightarrow r &< \alpha(1-p_0)^{n-1}((1-p_0) - n) \\ \Leftrightarrow \alpha &> \frac{r}{(1-p_0)^{n-1}((1-p_0) - n)}. \end{aligned} \tag{3.60}$$

24. The elasticity of a differentiable function $f(k)$ at k is defined as $\frac{\partial f}{\partial k}\frac{k}{f(k)}$.

Since $n > 1$, the right hand side is negative and the condition (3.60) is generally fulfilled. So, the condition of Theorem 1 is fulfilled. Since $u'(1) = 0$ for all $n > 1$, it follows from Theorem 1 that the upper belief cutoff is strictly smaller with remuneration $\bar{r} > 0$ than without.

Linear Belief Utility. [25] If $u(p) = \alpha(p-1)$, then, given (3.5), she chooses x at belief $p = \Pr(\alpha)$ if and only if

$$\begin{aligned} \alpha(p-1) + r &\geq -\alpha p \\ \Leftrightarrow 2\alpha p &\geq \alpha - r \\ \Leftrightarrow p &\geq \frac{1}{2} - \frac{r}{2\alpha}. \end{aligned} \tag{3.61}$$

She always prefers x regardless of her belief if

$$\begin{aligned} &\frac{1}{2} - \frac{r}{2\alpha} \leq 0 \\ &\Leftrightarrow \alpha \leq r. \end{aligned} \tag{3.62}$$

If (3.62) holds, it is optimal not to acquire any information and choose x. Conversely, when she is sufficiently altruistic, i.e. when $\alpha > r$, it is optimal to get fully informed about the state and choose the action that is harmless to the other in both states, i.e. x in X and y in Y.

25. The assumption of linear belief utility means that the agent is an expected utility maximizer.

References

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics* 115 (3): 715–753. [112]

Allison, Paul. 2002. "Bias in Fixed-Effects Cox Regression With Dummy Variables." *Manuscript, Department of Sociology, University of Pennsylvania*, [123]

Ambuehl, Sandro. 2017. "An Offer You Can't Refuse? Incentives Change What We Believe." *CESifo Working Paper Series No. 6296. Available at SSRN: https://ssrn.com/abstract=2917195*, [110]

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *Economic Journal* 100 (401): 464–477. [111]

Andreoni, James, and B Douglas Bernheim. 2009. "Social Image and the 50–50 Norm: A Theoretical and Experimental Analysis of Audience Effects." *Econometrica* 77 (5): 1607–1636. [111]

Andreoni, James, Justin M Rao, and Hannah Trachtman. 2017. "Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable giving." *Journal of Political Economy* 125 (3): 625–653. [110]

Ashton, Michael C, and Kibeom Lee. 2009. "The HEXACO–60: A short Measure of the Major Dimensions of Personality." *Journal of Personality Assessment* 91 (4): 340–345. [148, 149]

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka. 2016. "Attention Discrimination: Theory and Field experiments with Monitoring Information Acquisition." *American Economic Review* 106 (6): 1437–75. [110]

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–1678. [112]

Bénabou, Roland, and Jean Tirole. 2011. "Identity, Morals, and Taboos: Beliefs as Assets." *Quarterly Journal of Economics* 126 (2): 805–855. [111, 112]

Bénabou, Roland, and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30 (3): 141–64. [107, 110]

Blossfeld, Hans-Peter, Gotz Rohwer, and Thorsten Schneider. 2019. *Event History Analysis with Stata*. Routledge. [122]

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg registration and organization online tool." *European Economic Review* 71: 117–120. [118]

Bodner, Ronit, and Drazen Prelec. 2003. "Self-Signaling and Diagnostic Utility in Everyday Decision Making." *Psychology of Economic Decisions* 1: 105–26. [112]

Bolton, Gary E, and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (1): 166–193. [111]

Brier, Glenn W. 1950. "Verification of forecasts expressed in terms of probability." *Monthly weather review* 78 (1): 1–3. [123]

Brunnermeier, Markus K, Christian Gollier, and Jonathan A Parker. 2007. "Optimal Beliefs, Asset Prices, and the Preference for Skewed Returns." *American Economic Review* 97 (2): 159–165. [111]

Brunnermeier, Markus K, and Jonathan A Parker. 2005. "Optimal Expectations." *American Economic Review* 95 (4): 1092–1118. [111]

Charness, Gary, and Matthew Rabin. 2002. "Understanding Social Preferences with Simple Tests." *Quarterly Journal of Economics* 117 (3): 817–869. [111]

Cleves, M, W Gould, R Gutierrez, and Y Marchenko. 2010. *An Introduction to Survival Analysis Using Stata*. College Station, TX, Stata Press. [123]

Cox, David R. 1975. "Partial Likelihood." *Biometrika* 62 (2): 269–276. [146]

Dana, Jason, Roberto A Weber, and Jason Xi Kuang. 2007. "Exploiting Moral Wiggle Room: Experiments Demonstrating an Illusory Preference for Fairness." *Economic Theory* 33 (1): 67–80. [110, 111, 119, 127, 143, 152, 154]

DellaVigna, Stefano, John A List, and Ulrike Malmendier. 2012. "Testing for Altruism and Social Pressure in Charitable Giving." *Quarterly Journal of Economics* 127 (1): 1–56. [110]

Di Tella, Rafael, Ricardo Perez-Truglia, Andres Babino, and Mariano Sigman. 2015. "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism." *American Economic Review* 105 (11): 3416–42. [110]

Ditto, Peter H, and David F Lopez. 1992. "Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions." *Journal of Personality and Social Psychology* 63 (4): 568. [110]

Eil, David, and Justin M Rao. 2011. "The Good News-Bad News Effect: Asymmetric Processing of Objective Information About Yourself." *American Economic Journal: Microeconomics* 3 (2): 114–38. [110]

Exley, Christine, and Judd B Kessler. 2018. "Motivated Errors." [110]

Falk, Armin, Anke Becker, Thomas J Dohmen, David Huffman, and Uwe Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Netspar Discussion Paper*, [149]

Falk, Armin, and Nora Szech. 2016. "Pleasures of Skill and Moral Conduct." *CESifo Working Paper Series*, [110]

Fehr, Ernst, and Klaus M Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114 (3): 817–868. [111]

Feiler, Lauren. 2014. "Testing Models of Information Avoidance with Binary Choice Dictator Games." *Journal of Economic Psychology* 45: 253–267. [110, 143, 152, 154]

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental economics* 10 (2): 171–178. [118]

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological Games and Sequential Rationality." *Games and Economic Behavior* 1 (1): 60–79. [111]

Gino, Francesca, Michael I Norton, and Roberto A Weber. 2016. "Motivated Bayesians: Feeling Moral While Acting Egoistically." *Journal of Economic Perspectives* 30 (3): 189–212. [107]

Gneezy, Uri, Silvia Saccardo, Marta Serra-Garcia, and Roel van Veldhuizen. 2016. "Motivated Self-Deception, Identity and Unethical Behavior." *Working Paper*, [110]

Golman, Russell, David Hagmann, and George Loewenstein. 2017. "Information Avoidance." *Journal of Economic Literature* 55 (1): 96–135. [110]

Grossman, Zachary. 2014. "Strategic ignorance and the robustness of social preferences." *Management Science* 60 (11): 2659–2665. [110]

Grossman, Zachary, and Joël van der Weele. 2017. "Self-Image and Willful Ignorance in Social Decisions." *Journal of the European Economic Association* 15 (1): 173–217. [112]

Haisley, Emily C, and Roberto A Weber. 2010. "Self-Serving Interpretations of Ambiguity in Other-Regarding Behavior." *Games and Economic Behavior* 68 (2): 614–625. [110]

Kahneman, Daniel. 2003. "Experienced Utility and Objective Happiness: A Moment-Based Approach." In *The Psychology of Economic Decisions*. Vol. 1, Oxford: Oxford University Press, 187–208. [111]

Kahneman, Daniel, Peter P Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics* 112 (2): 375–406. [111]

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review* 101 (6): 2590–2615. [109, 111, 127, 129, 142]

Konow, James. 2000. "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions." *American economic review* 90 (4): 1072–1091. [111]

Köszegi, Botond. 2006. "Ego Utility, Overconfidence, and Task Choice." *Journal of the European Economic Association* 4 (4): 673–707. [111]

Lee, Kibeom, and Michael C Ashton. 2006. "Further Assessment of the HEXACO Personality Inventory: Two New Facet Scales and an Observer Report Form." *Psychological Assessment* 18 (2): 182. [149]

Lee, Kibeom, and Michael C Ashton. 2018. "Psychometric Properties of the HEXACO-100." *Assessment* 25 (5): 543–556. [117]

Loewenstein, George. 1987. "Anticipation and the Valuation of Delayed Consumption." *Economic Journal* 97 (387): 666–684. [111]

Masatlioglu, Yusufcan, A Yesim Orhun, and Collin Raymond. 2017. "Preferences for Non-Instrumental Information and Skewness." *Working Paper*, [110]

Mobius, Markus M, Muriel Niederle, Paul Niehaus, and Tanya S Rosenblat. 2011. "Managing Self-Confidence: Theory and Experimental Evidence." Working paper. National Bureau of Economic Research. [110]

Morris, Stephen, and Philipp Strack. 2019. "The Wald Problem and the Equivalence of Sequential Sampling and Static Information Costs." *Available at SSRN: https://ssrn.com/abstract=2991567 or http://dx.doi.org/10.2139/ssrn.2991567*, [131]

Moshagen, Morten, Benjamin E Hilbig, and Ingo Zettler. 2014. "Faktorenstruktur, Psychometrische Eigenschaften und Messinvarianz der Deutschsprachigen Version des 60-item HEXACO Persönlichkeitsinventars." *Diagnostica*, [149]

Murphy, Ryan O, Kurt A Ackermann, and Michel Handgraaf. 2011. "Measuring Social Value Orientation." *Judgment and Decision Making* 6 (2): 771–781. [117, 123, 148]

Niehaus, Paul. 2014. "A Theory of Good Intentions." *San Diego, CA: University of California and Cambridge, MA: NBER*, [111]

Rabin, Matthew. 1994. "Cognitive dissonance and social change." *Journal of Economic Behavior & Organization* 23 (2): 177–194. [111]

Rabin, Matthew. 1995. "Moral Preferences, Moral Constraints, and Self-Serving Biases." [111]

Raven, John Carlyle et al. 1998. *Raven's progressive matrices and vocabulary scales*. Oxford pyschologists Press. [117]

Schweizer, Nikolaus, and Nora Szech. 2018. "Optimal Revelation of Life-Changing Information." *Management Science* 64 (11): 5250–5262. [111]

Serra-Garcia, Marta, and Nora Szech. 2019. "The (in) Elasticity of Moral Ignorance." *CESifo Working Paper*, [110]

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69 (1): 99–118. [109, 131]

Spiekermann, Kai, and Arne Weiss. 2016. "Objective and Subjective Compliance: A Norm-Based Explanation of 'Moral Wiggle Room'." *Games and Economic Behavior* 96: 170–183. [111]

Zimmermann, Florian. Forthcoming. "The Dynamics of Motivated Beliefs." Working paper. [110]