Cognition, Argumentation, and Informed Choice*

Catherine Hafer† Dimitri Landa‡

PRELIMINARY, May 24, 2018

Abstract

How consequential are the limitations of citizens' reasoning and cognition for the quality of collective choice? To address this question, we develop a sequence of formal models focusing on a key aspect of individuals' reasoning ability relevant to politics: making inferences from arguments. In the models we analyze, a society is composed of a mix of cognitively different types of agents: agents who can make efficient use of the information entailed in arguments to which they may be exposed (the Bayesian ideal), and agents who, similar to Conan Doyle's Dr. Watson, have a cognitive bias that prevents them from drawing efficient indirect inferences. We consider the divergence of positions and judgments of these types of agents in relation to (1) the complexity of the argument space; (2) the distribution of these cognitive types in the society; (3) the incentives that a society with a given distribution of types creates for potential speakers to make informative arguments to the citizens; and (4) the conditions for sustaining the equilibria with the greatest possible degree of informative argumentation in the multi-speaker argumentation game. Our key results show that the differences between the posterior beliefs and choices of cognitive types tend to decrease with the greater strategic and argumentational complexity of the environment, and with the greater prevalence of cognitively bounded types.

*

†Department of Politics, NYU, e-mail: catherine.hafer@nyu.edu
‡Department of Politics, NYU, e-mail: dimitri.landa@nyu.edu

1 Introduction

Citizens' cognition and reasoning affect their induced preferences over policies and candidates and, in consequence, also the choices of political actors seeking to influence those preferences. Because the quality of democratic governance turns on citizens' ability to select the right kind of public officials and to hold them accountable for their choices, how citizens reason and what they know appears critically important for proper functioning of democratic institutions. Not surprisingly, scholars of political behavior have often voiced concerns over the systematic findings by cognitive and social psychologists pointing to dramatic limitations on citizens' knowledge and to departures in their reasoning and decision-making from the ideal of rational inference.

How consequential are these limitations? What implications do they have for the practice of democratic governance? How much should democratic theorists worry?

We develop a sequence of formal models to address these questions. Our analysis focuses on a key aspect of individuals' reasoning ability relevant to poitics: making inferences from arguments. To build a basic intuition underlying our treatment of arguments, note that one can distinguish three salient types of messages/arguments: messages that resonate with the receivers and, in so doing, either reinforce or directly persuade them of how they should think about the choices they face; messages that are simply not on point, and, as such can be ignored without loss; and messages that do not resonate but could, albeit indirectly and upon deeper reflection, reveal something valuable about the alternatives. A famous exchange between Sherlock Holmes and Dr Watson in Arthur Conan Doyle's "Silver Blaze,' illustrates the nature of the inference that is entailed in making sense of the latter kind of messages. Whereas an educated, but not particularly logically outstanding, Watson enirely disregards as irrelevant the fact that, in the story, the dog did not bark in the night, Holmes infers as elementary that it implies that the crime was not a burglary.

Holmes and Watson differ in their abilities to make an inference from the "null observation." In the context of argumentation, an equivalent of the "null observation" is the unpersuasive argument, provided that the listener knows something about the correlation between that argument and the position that it is meant to support. Another, more directly politically relevant example helps clarify some of the key incentives that may arise in relation to this point. Suppose that Jill is uncertain in her position on abortion rights: she leans against them, but is not yet absolutely convinced that that is the "right" position. Suppose, further, that Jill finds herself in a discussion with someone she has good reason to believe is one of the most thoughtful critics of abortion, and that this person makes a series of arguments, none of which Jill finds persuasive. How should this interaction affect her beliefs about the justifiability of abortion rights? If Jill is a Bayesian agent, she should conclude from the conjunction of two facts, her exposure to the most persuasive arguments possible against abortion rights and the unpersuasiveness to her of those arguments, that abortion rights are more defensible than she had previously thought. If Jill is, however, more like Watson in the Conan Doyle stories, she may be expected to fail to make that inference. If the critic of abortion rights that Jill met believes that Jill's position will change only if she hears (direct) arguments that she finds persuasive, then he loses nothing by making his best case to her. If, however, he believes that Jill, who would otherwise be likely to support anti-abortion politicians, would make inferences from the "null observation" (that is, from the fact that she has not heard a persuasive argument in his speech), he may be reluctant to make his arguments to her.

In the formal setting we describe below, we consider a society that is composed of a mix of Bayesian (Holmesian) and Watsonian types of agents and consider the divergence of positions and judgments of the two types of agents in relation to (1) complexity of the argument space; (2) the distribution of these cognitive types in the society; (3) the incentives that a society with a given distribution of types creates for potential speakers to address informative arguments to the citizens; and (4) the conditions for sustaining equilibria with greatest possible degree of argumentation in multiple-speaker argumentation game.

To summarize our key results: while greater cognitive sophistication (weakly) increases individual's welfare, its value added is fundamentally contingent on the informational and strategic properties of the interactive environment. Most starkly, we show that the social implications of greater cognitive sophistication may be negative, rather than positive: while Bayesians make better inferences, Watsonians encourage more speech, and the tendency of the net effect favors the society of Watsonians.

2 Relation to Literature

In modern political science, the tradition of concern about the implications of voters' cognition and reasoning for democratic governance originates with the seminal contributions by the Michigan School. If voters' information, attention, and inferential abilities are particularly limited, there may be good reasons to worry about the prospect of rational voting decisions, both in choosing the best policy in direct democracy settings, and in making the best evaluation of incumbent representative's performance in office.[1]

Lupia and McCubbins (1998) present a systematic critique of the Michigan School conclusions focusing on the assumption that individual knowledge of facts is necessary for high-quality choices. Rather than learning these facts themselves, they argue, citizens in democracies take cues from informed elites, and effective democratic institutions are those that enable most efficient identification of which elites to trust and for what cues.

The approach we take in this paper is complementary to theirs. Setting aside the cheap-talk and facts, we focus, rather, on arguments and cognition, and the general equilibrium implications of those for what kind of information is available and, ultimately, for the distribution of posterior beliefs in the society.

The formal literature on communication has focused primarily on the analysis of cheap-talk models of information transmission, in which the key question of interest is the conditions under which the senders are willing and able to reveal more or less coarse information to the receivers ((Crawford and Sobel 1982; Battaglini 2002; Meirowitz 2006; Gerardi and Yariv 2006; Austen-Smith and Feddersen 2006). In such models, the messages transmitted are "cheap" both in the sense of not costing anything to the sender and in that they are "merely" unverifiable "talk."

The approach closer to the present paper is that of the models of verifiable messages, in which senders can supply (partial) proofs for their signals and in which the *veridicality* (truth content) of a message is the same for all types of receivers (Lipman and Seppi 1995; Glazer and Rubinstein 2001, 2006; Landa and Meirowitz 2009; Mathis 2011). While the present model can be interpreted as one with verifiable (provable) messages,

[1]As a counterpoint, studies of aggregate voting decisions and attitudes (Erickson, Stimson, et al.) have painted a starkly contrasting view, largely consistent with the possibility of voters being efficient and effective principals and raising a puzzle of the apparent incompatibility with the voter-level studies.

the veridicality of messages we model differs across receivers, and when they reject the proof (that is, when the arguments are not convincing), the messages received may still have informational content that is either exogenously fixed or, as in cheap-talk models, endogenously derived from the equilibrium incentives of the players.

The closest models to the one developed in the present paper are Hafer and Landa (2007, 2008); Dickson, Hafer, and Landa (2008, 2015); and Hummel (2012). The first four papers analyze related models with argumentation that, like in the present model, satisfies full provability and private veridicality but unlike the current model's focus, their focus is on the informational effects of institutions affecting the speaker's access to the audience and of debate and preference aggregation rules, respectively. Hummel builds on the argumentation technology in the Hafer and Landa papers and introduces electoral competition with candidates taking positions before arguments regarding them are revealed to the voters, showing divergence in candidate platforms in two-candidate winner-take-all elections.

The non-Bayesian agency we model is closely connected to one of the central findings in cognitive psychology — the existence of a cognitive bias in favor of one's own currently held convictions (Zaller 1992; Dawes 1998; Rabin 1998; Baron 1994). An implication of such a bias is that in order to overturn one's prior position, the evidence against it or in favor of an alternative must be stronger than what be necessary if an update on that evidence was efficient. A direct and largely unambiguous evidence would stand a strongest chance for such an update, as agents "may even come to regard the ambiguities and conceptual flaws in the data opposing their hypotheses as somehow suggestive of the fundamental correctness of those hypotheses" (Lord et al. 1979, p. 2099). A closely related literature on the psychology of hypothesis testing (Wason 1968; 1977; Baron 1994, Ch. 13) suggests a similar idea: that people tend to look for "positive" confirmations of hypothesized patterns while disregarding or failing to look for "negative" information that does not fit the expected pattern but that can disconfirm the hypothesis.[2] Finally, and most directly relevant to the agency model in this paper are the experimental findings in Dickson, Hafer, and Landa (2007) on the consequences of the Watsoinian agency in

[2]There is also connection to the literature on overconfidence (in particular, Ortoleva and Snowberg (2015)), since overconfidence and confirmatory bias are close concepts.

a strategic communication setting. The authors report strong and robust evidence that a very substantial share of the subjects in the experiment behave consistent with our model of Watsonian subjects.

3 Model

Suppose a finite set N of (potential) speakers, $|N| = n$, such that their ideal points $\{\pi_1, \pi_2, ..., \pi_n\}$ are evenly spaced on the perimeter of a unit circle with center $(0, 0)$.[3] Throughout, we will refer to potential speakers as *speakers*, understanding by that term someone who can choose to send a message (equal to her position) or not, to the listeners. For all but the last part of the paper, we assume that there is a speaker associated with each of the N possible ideal points. We will use the term *active speaker* to refer to a speaker who does send a message (equal to her position) to the listeners. Let M be the set and m be the number of active speakers.

Initially, each listener does not know her ideal points $\pi_0 \in \{\pi_1, \pi_2, ..., \pi_n\}$ and assigns equal probability to elements in that set.[4] Thus, her initial expected value $E[\pi_0] = (0, 0)$. The listeners may learn their ideal points through speakers' communication. The exact nature of learning is explored below. After observing speakers' speeches (if any) and updating her beliefs, each listener chooses an action $\pi \in \mathbb{R}^2$, which we will refer to as that listener's policy choice.

We will say that an *argument i is persuasive to a listener j* when j directly learns her ideal point from hearing i, i.e., $i = j$. Otherwise, the argument i is unpersuasive to j – even though, it may be indirectly informative to j of her ideal point.

Let π_j represent speaker j's ideal point; also the true ideal point of a listener persuaded by j, $\pi_j = (\cos \frac{360}{n} j, \sin \frac{360}{n} j)$.

[3]We could have more than n speakers, with some speakers sharing the same ideal point, but if the cost of speaking $c > 0$, then in pure strategy equilibria, at most one speaker with a given ideal point will actually speak. So, for simplicity, we assume both that there are no duplicate speakers and no direct cost of speaking.

[4]In the present paper, we abstract away from the possibility that the listener may be persuaded by more than one distinct speaker, which would have an effect of allowing the listener's ideal point be in the interior of the circle. We leave the exploration of that possibility to future work.

The utility of speaker j when the listener chooses policy π is

$$-||\pi_j - \pi||^2.$$

We assume that listeners and speakers believe that each of the n arguments is equally likely to be persuasive to a given listener, but that only one of them can, in fact, be persuasive to that listener.

We will consider two cognitive types of listeners, to whom we will refer as *Bayesians* and *Watsonians*. The salient difference between these listeners is how they respond to unpersuasive arguments. To define it formally, suppose that they each hear argument j and find it unpersuasive. For the Bayesian listeners, the posterior p_B is:

$$\begin{aligned} \Pr(i \text{ is true}|\neg j) &= \frac{1}{n-1} \ \forall i \neq j \\ \Pr(j \text{ is true}|\neg j) &= 0. \end{aligned}$$

For Watsonian listeners, the posterior p_N is given by

$$\Pr(i \text{ is true}|\neg j) = \frac{1}{n} \ \forall i \text{ (inlcuding } i = j).$$

3.1 Bayesians vs. Watsonians: a Purely Informational Story

Our first observation is a trivial one: at any given profile of speeches, Bayesian listeners make (weakly) better policy choices on average than do Watsonians, and the relationship is strict if there is no active speaker at every possible ideal point. Our next question is about the distance between their posterior beliefs. We consider the divergence between these posteriors using a standard entropy-based divergence measure that does not require absolute continuity: *total variational distance* (also known as *statistical distance*).[5]

The *total variational distance* for posteriors p_B and p_N is

$$\frac{1}{2}\sum_{i=1}^{n}|p_B(i) - p_N(i)| = \frac{1}{2}\left(\frac{1}{n} + (n-1)\left(\frac{1}{n-1} - \frac{1}{n}\right)\right) = \frac{1}{n}.$$

Taking the limits, we get

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} &= 0 \\ \lim_{n\to 2} \frac{1}{n} &= \frac{1}{2}. \end{aligned}$$

[5]In the appendix, we show that another sometimes used measure that does not require absolute continuity, Lin's *symmetric divergence* measure, is comparable in our setting.

The more there is to learn (larger n), the smaller the divergence between the two cognitive types' beliefs when they are unpersuaded by a given argument. After a persuasive argument, total variational distance is 0 for either type of agent.

Suppose that agents learn m unpersuasive arguments, then

$$\frac{1}{2}\sum_{i=1}^{n}|p_B(i)-p_N(i)| = \frac{1}{2}\left(\frac{m}{n}+(n-m)\left(\frac{1}{n-m}-\frac{1}{n}\right)\right)=\frac{m}{n}$$

and

$$\begin{aligned}\lim_{n\to\infty}\frac{m}{n} &= 0\\ \lim_{m\to n-1}\frac{m}{n} &= \frac{n-1}{n}\\ \lim_{n\to\infty}\frac{n-1}{n} &= 1.\end{aligned}$$

Here, the more agents learn, the bigger the divergence between the cognitive types' beliefs when they remain unpersuaded by what they have heard.[6]

Summarizing the preceding conclusions, we have the following result:

Proposition 3.1. *Among listeners who have heard only unpersuasive arguments, the average differences in posterior beliefs between Bayesians and Watsonians are*

1. *greater when they have heard more arguments, holding fixed the number of possible arguments;*

2. *smaller when there are more possible arguments, holding fixed the number of arguments they have heard.*

The differences in beliefs between Bayesians and Watsonians are greater when they are exposed to more information, but less important when the issue is relatively more complicated, with more possible positions that can be supported by potentially compelling arguments.

[6] However, more agents are persuaded, and learn their ideal points with certainty, when they are exposed to more arguments.

3.2 The Incentives to Speak When There Are no Other Speakers

Next, we consider how the composition of the audience affects the speaker's incentives to speak when she is the only possible speaker.

Note that the ex ante expected value of the listener's ideal point is $(0,0)$. Suppose that argument j is unpersuasive. Denote by $\pi_{\neg j}$ the Bayesian's new (posterior) best policy choice following the unpersuasive argument. We have

$$\pi_{\neg j} = \sum_{i \neq j} \frac{1}{n-1} \left(\cos\left(\frac{360}{n} i\right), \sin\left(\frac{360}{n} i\right) \right).$$

Speaker j's ideal point is

$$\pi_j = \left(\cos\left(\frac{360}{n} j\right), \sin\left(\frac{360}{n} j\right) \right).$$

By the Pythagorean Theorem, the distance between π_j and $\pi_{\neg j}$ is $\frac{n}{n-1}$. Thus, j's expected utility from making argument j to a Bayesian listener is

$$\begin{aligned} E[u_j(j)] &= -\frac{1}{n}(0) - \frac{n-1}{n}\left(\frac{n}{n-1}\right)^2 \\ &= -\frac{n}{n-1}. \end{aligned}$$

The expected utility from not speaking is $E[u_j(\text{no speech})] = -1$. Thus:

$$E[u_j(\text{no speech})] - E[u_j(j)] = \frac{1}{n-1} > 0.$$

It is clear that

$$\lim_{n\to\infty} \left(E[u_j(\text{no speech})] - E[u_j(j)]\right) = 0,$$

but the covergence is "from above," that is, a single speaker facing a draw from a population of Bayesians will always prefer not speaking.

Facing a Watsonian listener, on the other hand, the lone speaker prefers speaking:

$$E[u_j(\text{no speech})] - E[u_j(j)] = -1 - \left(\frac{1}{n}(0) - \frac{n-1}{n}(1)\right) = -\frac{1}{n} < 0.$$

The presence of Watsonians, thus, increases the expected benefits of speaking, while decreasing the expected losses. So, the more likely the Watsonians, the sooner, as $n \to \infty$, the expected utility of speaking will exceed the expected utility of not speaking. Thus, the more likely the Watsonians, the more informed the posteriors of the listeners will be. Given uncertainty regarding the type of the audience, and letting β be the proportion of listeners who are Bayesians,

$$\begin{aligned} E[u_j(\text{no speech})] - E[u_j(j)] &= \beta\left(\frac{1}{n-1}\right) + (1-\beta)\left(-\frac{1}{n}\right) \\ &= \frac{1+2\beta n - n - \beta}{n^2 - n}. \end{aligned}$$

Simplifying, wee see that the speaker prefers to speak when

$$1 + 2\beta n \leq n + \beta. \tag{1}$$

Let β^* be the value of β such that for $\beta < \beta^*$, the expected value of speaking exceeds the expected value of silence for a lone speaker, and the reverse holds for $\beta \geq \beta^*$. Then, from (**??**), we get $\beta^* = \frac{n-1}{2n-1}$, with $\lim_{n\to\infty} \beta^* = \frac{1}{2}$ and $\lim_{n\to 2} \beta^* = \frac{1}{3}$.

We have the following result:

Proposition 3.2. *1. The greater n, the smaller the proportion of Watsonians necessary to support the speaker's incentive to make an argument; however,*

2. Even for arbitrarily large n, the speaker prefers to make his argument only if a majority of listeners are Watsonians.

3.3 Multiple Speakers

Our results thus far imply a tension. Because Bayesians use information more effectively, populations with higher proportions of Bayesians β enjoy higher average welfare because other things being equal more listeners are making better choices. However, incentives to speak are weaker in populations with a higher proportion of Bayesians, and listeners of both cognitive types need the information that speech provides in order to make better choices. To evaluate this tension, we must determine not only the effects of various features of the environment on speaker incentives, as is our focus above, but what sets of active speakers can be sustained as equilibrium behavior given that speakers choose whether to be active in speech-making or not.

If there is a speaker for every ideal point, then for every β, there exists an all-speaking equilibrium so that $M = N$. In this case, the informational welfare of the society is the same for all values of β. However, the existence of such an equilibrium is special: when not every ideal point has a (potential) speaker associated with it, all speakers being active cannot be supported as an equilibrium for all β, suggesting that the general relationship between informational welfare and β is more subtle. In this section, we seek to characterize that relationship, and so focus our analysis on the case when not every potential kind of speaker may be present.

We first examine the incentives for a potential speaker to speak in the presence of a set of another active speaker, and then generalize to arbitrary set of other active speakers. We then endogenize the set of active speakers and argue that the most stable distributions of active speakers will satisfy rotational symmetry, and, consequently, will minimize the social upside of Bayesian listeners.

3.3.1 2-speaker case

We next consider the equilibrium behavior in the 2-speaker case, studying a given potential speaker j's incentives to speak when there is one other active speaker, i. Normalizing j's position to $0°$, i.e., $\pi_j = (1, 0)$, we have the following result:

Proposition 3.3. *The two-speaker equilibrium can be sustained if and only if*

$$\beta \leq \beta^* = \frac{(n-1)(n-2)}{(n-1)^2 + 2(n-1)\cos\frac{360}{n}i + 1 + (n-1)(n-2)},$$

which binds less for smaller number of ideal point positions n and bigger distance between the speakers.

Proof. See Appendix. □

Note that β^* is decreasing with n and it is maximized when $\cos\frac{360}{n}i = -1$, that is, when i is exactly opposite j. Thus, the denser the distribution of ideal positions, and the closer the active speakers are to one another, the more Watsonians would be needed to support the speech of both speakers in equilibrium.

We next consider how the incentives to speak solo compare to the incentives to speak in duet. When speaking solo,

$$E[u_j(j \text{ speak})] - E(u_j(\text{no speech})] = -\beta\frac{1}{n-1} + (1-\beta)\frac{1}{n}. \quad (2)$$

Normalizing j to 0^0, and supposing active speaker i,

$$\begin{aligned} & E[u_j(j \text{ speak}|i \text{ speak})] - E(u_j(j \text{ silent}|i \text{ speak})] & (3) \\ = \quad & \frac{1}{n}(1-\beta) - \frac{\beta}{n(n-1)(n-2)}\left[(n-1)^2 + 2(n-1)\cos\frac{360}{n}i + 1\right] & (4) \end{aligned}$$

The presence of speaker i increases j's incentive to speak when the value of expression (**??**) is lower than the value of (**??**), i.e., when

$$-\beta\frac{1}{n-1} + (1-\beta)\frac{1}{n} < \frac{1}{n}(1-\beta) - \frac{\beta}{n(n-1)(n-2)}\left[(n-1)^2 + 2(n-1)\cos\frac{360}{n}i + 1\right],$$

which reduces to

$$-\frac{1}{n-1} > \cos\frac{360}{n}i.$$

Note that this condition, which is independent of β, is easier to meet the more negative $\cos\frac{360}{n}i$ is – that is, the farther away from j it is and the greater the number of ideal points n there are.

3.3.2 Equilibrium with m speakers

Suppose, a set M of active speakers, recalling that $|M| = m$, and $M \subset N$, and consider how their presence affects the incentives to speak for another speaker $j \notin M$.

We begin with the following result, which creates a baseline expectation about the relationship between listener cognition and the potential speakers' incentives to speak:

Proposition 3.4. *Given a set of active speakers M, $m = |M| < n-2$, j's benefit from speaking is decreasing in the share of listeners who are Bayesian, β.*

Proof. See Appendix. □

Next, we have the following result:

Proposition 3.5. *Holding fixed the set of other active speakers M, j's benefit from speaking is*

1. *greater when the other active speakers' positions are farther from π_j; and*
2. *greater when the other active speakers' positions are symmetric around the bisection of the circle through π_j.*

Proof. See Appendix. $\square$

The following definition formalizes the idea of an equilibrium set of speakers:

Definition 3.1. *A set of active speakers M can be supported in equilibrium given β if for all $j \in M$, at that value of β*

$$E[u_j(j)|M\backslash\{j\}] - E[u_j(\textit{no speech})|M\backslash\{j\}] \geq 0$$

and for all $j \in N\backslash M$,

$$E[u_j(j)|M] - E[u_j(\textit{no speech})|M] \leq 0.$$

From Propositions **??** and **??**, the sets of active speakers that can be supported in equilibrium depend on both β and the configuration of their positions. While a dynamic model of speaker entry and exist is beyond the scope of this paper, we can examine the features of the equilibria that present marginal speakers and non-speakers with the strongest incentives to speak and be silent, respectively.

Let $\Delta(j, M; N, \beta) := E[u_j(\text{speaking}|M)] - E[u_j(\text{silent}|M\backslash\{j\})]$ be j's expected benefit of speaking in the presence of active speakers $M\backslash\{j\}$. The least motivated speakers in M are then $L(M; N, \beta) := \{l \in M : \Delta(l, M; N, \beta) \leq \Delta(k, M; N, \beta)\ \forall k \in M\}$.

Definition 3.2. *Let M', M'' each be a set of active speakers that can be supported in equilibrium s.t. $|M'| = |M''| = m$. The set of speakers M'' is* ***weakly more robust*** *than the set M' if, for $l'' \in L(M''; N, \beta)$ and $l' \in L(M'; N, \beta)$, $\Delta(l'', M''; N, \beta) \geq \Delta(l', M'; N, \beta)$, and* ***strictly more robust*** *if $\Delta(l'', M''; N, \beta) > \Delta(l', M'; N, \beta)$. Given the set of sets of active speakers that can be supported in equilibrium $\mathcal{M} = \{M', M'', M''', ...\}$ s.t. $|M'| = |M''| = ... = m$, the set $M \in \mathcal{M}$ is* ***maximally robust*** *if there exists no set $\check{M} \in \mathcal{M}$ that is strictly more robust than M.*

Let π_M be the expected value of the ideal points of the prior distribution of listeners over the ideal points of the set of active speakers M. We next state the following proposition:

Proposition 3.6. *(1) Let M', M'' each be a set of active speakers that can be supported in equilibrium, s.t. $|M'| = |M''| = m$. Then set of speakers M'' is more robust than the* *set M' if and only if $||\pi_0, \pi_{M''}|| < ||\pi_0, \pi_{M'}||$; and (2) The set M is maximally robust if and only if M is rotationally symmetric.*

Proof. See Appendix. $\square$

Let $m(\beta, S)$ be the largest number of active speakers supported in equilibrium for β, given S, and define $\mathbb{M}(\beta, S) := \{M \subseteq S : |M| = m(\beta, S)$ and M can be supported in equilibrium$\}$. Let $\mathbb{M}^*(\beta, S)$ be the set of maximally robust M in $\mathbb{M}(\beta, S)$, $\mathbb{M}^*(\beta, S) \subseteq \mathbb{M}(\beta, S)$. Let $M^* \in \mathbb{M}^*(\beta, S)$ and $l^* \in L(M^*, N, \beta)$.

Define $\beta^*(M^*)$ to be the value of β such that l^* is indifferent between speaking and not, given M:

$$\begin{aligned}\beta^*(M^*) \;:&= \left(1 + \frac{1}{(n-m+1)(n-m+2)}\left[\left((n-m)\cos\frac{360}{n}l^* + \sum_{i \in M^*}\cos\frac{360}{n}i\right)^2\right.\right.\\ &\left.\left.+\left((n-m)\sin\frac{360}{n}l^* + \sum_{i \in M}\frac{1}{n-s}\sin\frac{360}{n}i\right)^2\right]\right)^{-1}\end{aligned}$$

Thus, $\beta^*(M^*)$ is the upper bound on the proportion of Bayesian listeners such that the largest possible number of active speakers can be supported in equilibrium.

The following two results characterize the equilibrium properties when the set of potential speakers is rotationally symmetric.

Proposition 3.7. *Let $S \subset N$ be rotationally symmetric, and $n - s \geq 2$. Then for $\beta \leq \beta^*(M^*) = \frac{n-s-1}{2(n-s-1)-1}$, $M^* = S$ can be supported in equilibrium, and for $\beta > \beta^*(M^*)$, there are no active speakers in equilibrium.*

Proof. See Appendix. $\square$

In Appendix B, we next consider the properties of equilibria when the set of potential speakers S is not rotationally symmetric, but at least some subset of it is. While the characterization there is somewhat more subtle, the social welfare implication from Proposition **??**, though somewhat augmented, remains a feature of the equilibrium outcome.

3.3.3 Welfare Analysis

We provide two results on the effects of model parameters on the listener utility.

Proposition 3.8. *Let $S \subset N$ be rotationally symmetric, and $n - s \geq 2$. Then for $M^* = S$, average listener utility in equilibrium is constant in β for $\beta < \beta^*(M^*) = \frac{n-s-1}{2(n-s-1)-1}$, drops discontinuously to -1 at $\beta = \beta^*(M^*)$, and is constant in β at -1 for $\beta > \beta^*(M^*)$.*

Proof. See Appendix. ☐

Proposition **??** suggests a stark conclusion. When the set of potential speakers is rotationally symmetric, the net value of the tradeoff between better inferences (from higher proportion of Bayesians) and encouraging more speech (from higher proportion of Watsonians) is such that the society is better off with Watsonians than Bayesians. If we think of utility as a measure of agent fitness, the individual-level premium on developing hyper-rationality (efficient, Holmesian learning) may be limited.

Here is another proposition:

Proposition 3.9. *(1) The difference in the equilibrium expected utility of Bayesian and Watsonian listeners decreases as $||\pi_0, \pi_M||$ decreases. (2) If M is rotationally symmetric, then Bayesians and Watsonians listeners make the same policy choices in equilibrium and obtain the same equilibrium payoffs.*

Proof. See Appendix. ☐

If M is not rotationally symmetric, then for β sufficiently low, individual listeners are better off being Bayesian because there is information available but making use of it requires efficient inference. But once β is high enough, individual listeners are indifferent between being Bayesian and Watsonian because the speakers are not providing information.

When the active speakers are rotationally symmetric – that is, when the set of active speakers is maximally robust – the policy choices of unpersuaded Bayesians are the same as those of the unpersuaded Watsonians. Thus, there are no equilibrium benefits of greater cognitive ability, and the benefits of having more speakers (associated with lower β) dominate.

4 Conclusion

To be Added.

5 Appendix

Lin's *symmetric divergence* measure for probability distributions p_1 and p_2 of a discrete random variable X, $L(p_1, p_2)$, is a sum of two directed (*asymmetric*) divergence measures $K(p_1, p_2)$ and $K(p_1, p_2)$, where $K(p_i, p_j)$ is defined as

$$K(p_i, p_j) = \sum_{x \in X} p_i(x) \log_2 \left(\frac{p_i(x)}{\frac{1}{2}p_i(x) + \frac{1}{2}p_j(x)} \right).$$

Suppose that agents heard one unpersuasive argument. Then

$$\begin{aligned}
K(p_B, p_N) &= 0 + (n-1)\frac{1}{n-1} \log_2 \left(\frac{\frac{1}{n-1}}{\frac{1}{2}\frac{1}{n-1} + \frac{1}{2}\frac{1}{n}} \right) = \log_2 \left(\frac{2n}{2n-1} \right) \\
K(p_N, p_B) &= \frac{1}{n} \log_2 2 + (n-1)\frac{1}{n} \log_2 \left(\frac{\frac{1}{n}}{\frac{1}{2}\frac{1}{n} + \frac{1}{2}\frac{1}{n-1}} \right) = \frac{1}{n}(1 + (n-1) \log_2 \left(\frac{2(n-1)}{2n-1} \right) \\
L(p_B, p_N) &= \log_2 \left(\frac{2n}{2n-1} \right) + \frac{1}{n} + \frac{n-1}{n} \log_2 \left(\frac{2(n-1)}{2n-1} \right).
\end{aligned}$$

Further,

$$\begin{aligned}
\lim_{n \to 2} L(p_B, p_N | n &= 2) = 3 - \frac{3}{2} \log_2 3 \approx 0.622\,56 \\
\lim_{n \to \infty} L(p_B, p_N) &= 0.
\end{aligned}$$

Proof of Proposition ??

If a listener is a Watsonian, then her posterior remains $(0, 0)$ if she is unpersuaded, and moves to her her ideal point equal to the speech if she is persuaded. If a listener is Bayesian, then if she is unpersuaded by i alone, then

$$\pi_{\neg i} = \frac{1}{n-1} \left(-\cos \frac{360}{n} i, -\sin \frac{360}{n} i \right),$$

and if she is unpersuaded by both i and and arbitrary j, then

$$\pi_{\neg i \neg j} = \frac{1}{n-2} \left(-\cos \frac{360}{n} i - \cos \frac{360}{n} j, -\sin \frac{360}{n} i - \sin \frac{360}{n} j \right).$$

The incentive to speak to a Watsonians is $\frac{1}{n}(1-\beta) > 0$. To see the incentive to speak to Bayesians is (restricting attention to Bayesian audience):

$$
\begin{aligned}
& E[u_j(j \text{ speak}|i \text{ speak})] - E(u_j(j \text{ silent}|i \text{ speak})] \\
= & \frac{n-2}{n}\left(||\pi_j, \pi_{\neg i}||^2 - ||\pi_j, \pi_{\neg i \neg j}||^2\right) + \frac{1}{n}(0) + \frac{1}{n}(0 + ||\pi_j, \pi_{\neg i}||^2) \\
= & \frac{n-1}{n}||\pi_j, \pi_{\neg i}||^2 - \frac{n-2}{n}||\pi_j, \pi_{\neg i \neg j}||^2 \\
= & \frac{n-1}{n}\left(\left(\cos\frac{360}{n}j + \frac{1}{n-1}\cos\frac{360}{n}i\right)^2 + \left(\sin\frac{360}{n}j + \frac{1}{n-1}\sin\frac{360}{n}i\right)^2\right) \\
- & \frac{n-2}{n}\left(\left(\cos\frac{360}{n}j + \frac{1}{n-2}\cos\frac{360}{n}i + \frac{1}{n-2}\cos\frac{360}{n}j\right)^2\right. \\
+ & \left.\left(\sin\frac{360}{n}j + \frac{1}{n-2}\sin\frac{360}{n}i + \frac{1}{n-2}\sin\frac{360}{n}j\right)^2\right)
\end{aligned} \tag{5}
$$

Normalizing j to 0, this reduces to

$$
\begin{aligned}
= \quad & \frac{n-1}{n}\left(\left(1 + \frac{1}{n-1}\cos\frac{360}{n}i\right)^2 + \left(\frac{1}{n-1}\sin\frac{360}{n}i\right)^2\right) \\
& -\frac{n-2}{n}\left(\left(\frac{n-1}{n-2} + \frac{1}{n-2}\cos\frac{360}{n}i\right)^2 + \left(\frac{1}{n-2}\sin\frac{360}{n}i\right)^2\right),
\end{aligned}
$$

which, after some algebra, reduces to

$$
-\frac{1}{n(n-1)(n-2)}\left[(n-1)^2 + 2(n-1)\cos\frac{360}{n}i + 1\right] < 0.
$$

Note that this difference is larger (less negative) for i farther from j, maximizing at $180^{0.}$.

To support j's speech in the presence of i's speech, we need

$$
\frac{1}{n}(1-\beta) - \frac{\beta}{n(n-1)(n-2)}\left[(n-1)^2 + 2(n-1)\cos\frac{360}{n}i + 1\right] \geq 0, \tag{6}
$$

Solving for the critical value of β, we get

$$
\beta^* = \frac{(n-1)(n-2)}{(n-1)^2 + 2(n-1)\cos\frac{360}{n}i + 1 + (n-1)(n-2)},
$$

which is less than 1 for all i if $n \geq 3$. Thus, for all $\beta \leq \beta^*$, speaker i will prefer to speak if speaker j is speaking.

By observation, β^* is decreasing with n and it is maximized when $\cos\frac{360}{n}i = -1$, that is, when i is exactly opposite j. ■

Proof of Proposition ??

If all m speeches of the set M of speakers are unpersuasive, then the Bayesian listener's policy choice is

$$\pi_{\neg M} := \frac{1}{n-m}\left(-\sum_{i\in M}\cos\frac{360}{n}i, -\sum_{i\in M}\sin\frac{360}{n}i\right).$$

If j also speaks and is also unpersuasive, the listener's policy choice is

$$\frac{1}{n-m-1}\left(-\sum_{i\in M}\cos\frac{360}{n}i-\cos\frac{360}{n}j, -\sum_{i\in M}\sin\frac{360}{n}i-\sin\frac{360}{n}j\right).$$

To get the expected difference between speaking and not speaking for j, given the other M active speakers, we have to calculate $E[u_j]$, recognizing that if he speaks, then with probability $\frac{1}{n}$, he moves the listener to his position and, with probability $\frac{n-m-1}{n}$, the listener is unpersuaded by all $M \cup \{j\}$. First, suppose the listener is Bayesian. Then:

$$\begin{aligned}
& E[u_j(\text{no speech})|\text{all } i\in M \text{ speak}] - E[u_j(j)|\text{all } i\in M \text{ speak}] \\
= \; & -\frac{n-m}{n}\left[\left(\cos\frac{360}{n}j+\frac{1}{n-m}\sum_{i\in M}\cos\frac{360}{n}i\right)^2+\left(\sin\frac{360}{n}j+\frac{1}{n-m}\sum_{i\in M}\sin\frac{360}{n}i\right)^2\right] \\
& +\frac{n-m-1}{n}\left[\left(\cos\frac{360}{n}j+\frac{1}{n-m-1}\sum_{i\in M}\cos\frac{360}{n}i+\frac{1}{n-m-1}\cos\frac{360}{n}j\right)^2\right. \\
& \left.+\left(\sin\frac{360}{n}j+\frac{1}{n-m-1}\sum_{i\in M}\sin\frac{360}{n}i+\frac{1}{n-m-1}\sin\frac{360}{n}j\right)^2\right]+\frac{1}{n}(0)^2 \\
= \; & \frac{1}{n(n-m)(n-m-1)}\left[\left((n-m)\cos\frac{360}{n}j+\sum_{i\in M}\cos\frac{360}{n}i\right)^2+\left((n-m)\sin\frac{360}{n}j+\sum_{i\in M}\sin\frac{360}{n}i\right)^2\right]
\end{aligned}$$

Suppose, as before, that β listeners are Bayesians. For the $(1-\beta)$ Watsonian listeners, the difference for j between speaking to them and not is $\frac{1}{n}$.

Let $\Delta(j, M; N, \beta) \equiv E[u_j(j)|M] - E[u_j(\text{no speech})|M]$.

$$\begin{aligned}
&\Delta(j, M; N, \beta) && (7)\\
= \;& (1-\beta)\left(\frac{1}{n}\right) - \beta\left(\frac{1}{n}\right)\left(\frac{1}{(n-m)(n-m-1)}\right)\left((n-m)\pi_j + \sum_{i\in M}\pi_i\right)^2 \\
= \;& (1-\beta)\left(\frac{1}{n}\right) - \beta\frac{1}{n(n-m)(n-m-1)} \\
& \left[\left((n-m)\cos\frac{360}{n}j + \sum_{i\in M}\cos\frac{360}{n}i\right)^2 + \left((n-m)\sin\frac{360}{n}j + \sum_{i\in M}\sin\frac{360}{n}i\right)^2\right].
\end{aligned}$$

j prefers to speak in the presence of active speakers M if $\Delta(j, M; N, \beta) > 0$. Given that $(1-\beta)\left(\frac{1}{n}\right)$ is decreasing in β, and the remaining term in (**??**) is decreasing in β as well, the claim of the proposition follows. ■

Proof of Proposition ??

j's incentives to speak are greater when the constraint on β, β^*, is higher. Re-arranging (**??**), we get an equivalent condition for j's speech:

$$\begin{aligned}
&\beta \le \beta^* \\
= \;& \left\{1 + \frac{n-m}{n-m+1}\left[\left(\cos\frac{360}{n}j + \frac{1}{n-m}\sum_{i\in M}\cos\frac{360}{n}i\right)^2 + \left(\sin\frac{360}{n}j + \frac{1}{n-m}\sum_{i\in M}\sin\frac{360}{n}i\right)^2\right]\right\}^{-1} && (8)
\end{aligned}$$

Normalizing j to $0°$ for the ease of exposition,

$$\beta^* = \left\{1 + \frac{n-m}{n-m+1}\left[\left(1 + \frac{1}{n-m}\sum_{i\in M}\cos\frac{360}{n}i\right)^2 + \left(\frac{1}{n-m}\sum_{i\in M}\sin\frac{360}{n}i\right)^2\right]\right\}^{-1} \qquad (9)$$

β^* is higher when $\left(1 + \frac{1}{n-m}\sum_{i\in M}\cos\frac{360}{n}i\right)^2$ and $\left(\frac{1}{n-m}\sum_{i\in M}\sin\frac{360}{n}i\right)^2$ are smaller. The former is minimized when when $\sum_{i\in M}\cos\frac{360}{n}i$ is minimized, which corresponds to $\cos\frac{360}{n}i$ being as close to -1 as possible for as many $i \in M$ as possible. Given j is normalized to $\pi_j = 0°$, part (1) of the proposition follows. The latter term is minimized when $\forall i \in M$, there is a $k \in M$ s.t. $\sin\pi_k = -\sin\pi_i$, i.e., $\forall i \in M$ s.t. $\pi_i \in \{0°, 180°\}$, there is a $k \in M$ s.t. $\pi_k = 360° - \pi_i$. This establishes part (2). ■

5.1 Proof of Proposition ??

1. Consider $l' \in L(M'; N, \beta)$ and $l'' \in L(M''; N, \beta)$. From the fact that M' and M'' each can be supported in equilibrium, and $l \in L(M; N, \beta)$ only if $l \in M$, we have that $\Delta(l', M'; N, \beta) \geq 0$ and $\Delta(l'', M''; N, \beta) > 0$. For ease of comparison, we can normalize the choice of axes in M' so that l' is 0, and in M'' so that l'' is 0. Given that $|M'| = |M''| = m$, we see from the expression (??) for $\Delta(j, M; N, \beta)$, that $\Delta(l'', M''; N, \beta) > \Delta(l', M'; N, \beta)$ if and only if

$$\left((n-m)(1,0) + \sum_{i \in M''} \pi_i\right)^2 < \left((n-m)\pi_j + \sum_{i \in M'} \pi_i\right)^2,$$

or, equivalently, iff $\left|\sum_{i \in M''} \pi_i\right| < \left|\sum_{i \in M'} \pi_i\right|$, or $|\pi_{M''}| < |\pi_{M'}|$. The result, then, follows from $\pi_0 = (0, 0)$.

2. Follows immediately.

Proof of Proposition ??

Let $S \subset N$ be rotationally symmetric. From (??), $j \in S$ prefers to speak if

$$\begin{aligned} &(1-\beta)\left(\frac{1}{n}\right) - \beta \frac{1}{n(n-s)(n-s-2)} \\ &\left[\left((n-s-2)\cos\frac{360}{n}j + \sum_{i \in S} \cos\frac{360}{n}i\right)^2 + \left((n-s-2)\sin\frac{360}{n}j + \sum_{i \in S} \sin\frac{360}{n}i\right)^2\right] \\ \geq\ & 0. \end{aligned} \tag{10}$$

The rotational symmetry of S implies that $\sum\limits_{i \in S} \cos \frac{360}{n} i = 0$ and $\sum\limits_{i \in S} \sin \frac{360}{n} i = 0$. Making this substituution, simplifying, and isolating β yields $\beta \leq \beta^*(M^*) = \frac{n-s-1}{2(n-s-1)-1}$.

Note that $\beta^*(M^*)$ is increasing in s as s increases toward $n-2$. Thus, a larger rotationally symmetric $M \subseteq S$ can be supported over a larger range of β than can a smaller rotationally symmetric $\hat{M} \subset S$.

Now consider $M \subset S$, $\hat{M} \subset S$ s.t. $|M| = |\hat{M}|$, $j \in M \cap \hat{M}$, $\hat{M}$ is rotationally symmetric and M is not. Then inequality (??) can be satisfied at higher values of β for $\hat{M}$ than for M. Thus, the set of values for which any $M \subset S$ can be supported in equilibrium is a strict subset of the set of values $[0, \beta^*(M^*)]$ for which $M = S$ can be supported. For $\beta > \beta^*(M^*)$, $M = \emptyset$. ∎

Proof of Proposition ??

From the previous result, $\forall \beta < \beta^*(M^*)$, $M^* = S$. The average expected utility of a listener is, then,

$$-\frac{m}{n}(0) - \frac{n-m}{n}\left[(1-\beta)(1) + \beta\frac{1}{n-m}\sum_{i\in N\setminus M^*}\left(\left(\cos\frac{360}{n}i + \sum_{k\in M^*}\cos\frac{360}{n}k\right)^2 + \left(\sin\frac{360}{n}i + \sum_{k\in M^*}\sin\frac{360}{n}k\right)^2\right)\right]$$

$\forall \beta < \beta^*(M^*)$.

Because $M^* = S$ is rotationally symmetric, $\sum_{k\in M^*}\cos\frac{360}{n}k = 0$ and $\sum_{k\in M^*}\sin\frac{360}{n}k = 0$, the average listener utility can be re-written as

$$-\frac{n-m}{n}\left[(1-\beta)(1) + \beta\frac{1}{n-m}\sum_{i\in N\setminus M^*}(1)\right] = -\frac{n-m}{n},$$

which is constant in β.For $\beta > \beta^*(M^*)$, $M = \emptyset$. Because there is no speech, all listeners choose $\pi = \pi_0$ and receive utility of -1 $\forall \beta > \beta^*(M^*)$. ■

5.2 Proof of Proposition ??

1. Both Bayesians and Watsonians who earn their ideal points obtain the same payoffs, thus, we mat restrict attention to the expected payoffs of unpersuaded listeners.

An unpersuaded Watsonian chooses

$$\pi = \pi_0 = (0,0)$$

and obtains an expected payoff of

$$-\frac{1}{n-m}\sum_{i\in N\setminus M}||\pi_i,\pi_0||^2 = -1.$$

A Bayesian who is unpersuaded after M chooses

$$\pi = \frac{1}{n-m}\sum_{i\in N\setminus M}\pi_i = \frac{1}{n-m}\left(\sum_{i\in N}\pi_i - \sum_{i\in M}\pi_i\right) = \frac{1}{n-m}(n\pi_0 - m\pi_M) = \frac{1}{n-m}(-m\pi_M)$$

and obtains an expected payoff

$$-\frac{1}{n-m}\sum_{i\in N\setminus M}||\pi_i, -\frac{m}{n-m}\pi_M||^2.$$

Using the Pythagorian Theorem, and recalling that the radius of the circle is 1, this expected payoff can be re-written as

$$\begin{aligned}
& -\frac{1}{n-m}\sum_{i\in N\setminus M}\left[\left(\cos\frac{360}{n}i+\frac{1}{n-m}\sum_{j\in M}\cos\frac{360}{n}j\right)^2+\left(\sin\frac{360}{n}i+\frac{1}{n-m}\sum_{j\in M}\sin\frac{360}{n}j\right)^2\right] \\
= & -1+\frac{1}{(n-m)^2}\left[\left(\sum_{j\in M}\cos\frac{360}{n}j\right)^2+\left(\sum_{j\in M}\sin\frac{360}{n}j\right)^2\right] \\
= & -1+\frac{1}{(n-m)^2}||\pi_0, m\pi_M||^2.
\end{aligned}$$

Comparing the expected utilities of unpersuaded Bayesians abd unpersuaded Watsonians, we see that Bayesians have higher expected utility, that the difference is increasing in the distance between π_0 and π_M, and that when this distance goes to 0, the difference in expected utility also goes to 0.

Appendix B

Proposition 5.1. *Let $S \subset N$ and $n-2 \geq 2$. Suppose that S is not rotationally symmetric but some $S' \subset S$ is. Let $\hat{S}$ be the largest rotationally symmetric subset of S. Then:*

1. *$m(\beta, S)$ is weakly decreasing in β, with $m(0, S) = s$ and $m(1, S) = 0$.*
2. *For $\beta \in (\beta^*_{\hat{s}+1}, \beta^*_{\hat{s}})$, $\mathbb{M}^*(\beta, S) = \{\hat{S}\}$, and average listener utility is constant and greater than -1; for $\beta > \beta^*_{\hat{s}}$, $\mathbb{M}^*(\beta, S) = \{\emptyset\}$, and average listener utility is constant at -1. For $\beta < \beta^*_{\hat{s}+1}$, Bayesian and Watsonian listeners have the same payoffs.*
3. *The interval $[0, \beta^*_{\hat{s}+1})$ is partitioned such that on each $(\beta^*_{m+1}, \beta^*_m)$, $M^*(\beta, S)$ is not rotationally symmetric, average listener utility is increaasing in β on $(\beta^*_{m+1}, \beta^*_m)$, and Bayesian listeners have higher average utility than Watsonian listeners.*

Proof. Let $S \subset N$ and $\hat{S}$ satisfy the conditions in the statement of the proposition, and let $\hat{s} = |\hat{S}|$. Note that $M^*(0, S) = S$. Thus, $\exists \beta^*_m < \beta^*_{\hat{s}}$ for some $m > \hat{s}$. From

Propositions 5 and 6, $\forall \beta \in (\beta^*_{\hat{s}+1}, \beta^*_{\hat{s}})$, $M^*(\beta, S) = \hat{S}$ and average listener utility is constant at $-\frac{n-\hat{s}}{n}$. $\forall \beta > \beta^*_{\hat{s}}$, $M^*(\beta, S) = \emptyset$ and listener utility is constant at -1. For $\beta > \beta^*_{\hat{s}+1}$, $\pi_{\neg M^*(\beta,S)} = \pi_0$ and Bayesian and Watsonian listeners have the same expected utilities.

Now consider some $m \in \{\hat{s}+1, ..., s\}$. By definition, the largest rotationally symmetric subset of S is $\hat{S}$, so for $m > \hat{s}$, $M^*(\beta, S)$ associated with $\beta \in [\beta^*_{m+1}, \beta^*_m)$ is not rotationally symmetric. Because it is not rotationally symmetric, $\pi_{\neg M^*(\beta,S)} \neq \pi_0$, and because $\pi_{\neg M^*(\beta,S)}$ maximizes $-\frac{1}{n-m} \sum_{i \in N \backslash M^*(\beta,S)} ||\pi_i, \pi||^2$, it must be that this quantity is greater that -1 when evaluated at $\pi = \pi_{\neg M^*(\beta,S)}$. Thus, average listener utility

$$\frac{m}{n}(0) + \frac{n-m}{n}\left[(1-\beta)(-1) - \beta\frac{1}{n-m}\sum_{i \in N \backslash M^*(\beta,S)} ||\pi_i, \pi_{\neg M^*(\beta,S)}||^2\right]$$

is increasing in β on $[\beta^*_{m+1}, \beta^*_m)$.

Because $m(\beta, S)$ drops discontinuously at β^*_m $\forall m \in \{\hat{s}, ..., s\}$, it causes a downward jump discontinuity in average listener utility at β^*_m because

$$0 > -(1-\beta) - \beta\frac{1}{n-m}\sum_{i \in N \backslash M^*(\beta,S)} ||\pi_i, \pi_{\neg M^*(\beta,S)}||^2.$$

■

References

Austen-Smith, David, and Timothy Feddersen, 2006. "Deliberation, Preference Uncertainty, and Voting Rules." *American Political Science Review* 100 (2), 209-217.

Baron, Jonathan. 1994. *Thinking and Deciding.* Cambridge University Press.

Battaglini, Marco. 2002. "Multiple Referrals and Multidimensional Cheap-talk." *Econometrica* 70 (4), 1379-1401.

Calvert, Randall, L. 1985. "The Value of Biased Information: A Rational Choice Model of Political Advice." *Journal of Politics* 47 (2), 530-555.

Crawford, Vincent and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50, 1431-51.

Dewatripont, Mathias and Jean Tirole. 1999. "Advocates." *The Journal of Political Economy* 107, 1-39.

Dickson, Eric S., Catherine Hafer, and Dimitri Landa. 2008. "Cognition and Strategy: a Deliberation Experiment." *Journal of Politics* 70 (4), 974-989.

Dawes, Robyn M. 1998. "Behavioral Decision Making and Judgment." In D. T. Gilbert, et al. eds., *The Handbook of Social Psychology*, 4th ed. McGraw Hill, 497-548.

Gerardi, Dino and Leeat Yariv. 2006. "Deliberative Voting." *Journal of Economic Theory* 134, 317-338..

Glazer, Jacob and Ariel Rubinstein. 2001. "Debates and Decisions: on a Rationale for Argumentation Rules." *Games and Economic Behavior* 36, 158-73.

Glazer, Jacob and Ariel Rubinstein. 2006. "A Study in the Pragmatics of Persuasion: a Game Theoretical Approach." *Theoretical Economics* 1 (4), 395-410.

Hafer, Catherine and Dimitri Landa. 2007. "Deliberation as Self-Discovery and Institutions for Political Speech." *Journal of Theoretical Politics* 18 (3), 329-360.

Hafer, Catherine and Dimitri Landa. 2008. "Majoritarian Debate." New York University Mimeo.

Hummel, Patrick. 2012. "Deliberative Democracy and Electoral Competition." *Games and Economic Behavior* 75 (2), 646–667.

Landa, Dimitri and Adam Meirowitz. 2009. "Game Theory, Information, and Deliberative Democracy." *American Journal of Political Science* 53 (2), 427–444.

Lin, Jianhua. 1991. "Divergence Measures Based on the Shannon Entropy." *IEEE Transactions on Information Theory*, vol. 37, no. 1: 145-151.

Lipman, Barton and Duane J. Seppi. 1995. "Robust Inference in Communication Games with Partial Provability." *Journal of Economic Theory* 66 (2), 370-405.

Lord, C. G., L. Ross, and M. R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* XXXVII, 2098-2109.

Lupia, Arthur, and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

MacKuen, Michael B., Robert S. Erikson and James A. Stimson. 1992. "Peasants or Bankers? The American Electorate and the U.S. Economy." *American Political Science Review* 86 (3), 597-611.

Mathis, Jerome. 2011. "Deliberation with Evidence." *American Political Science Review* 105 (3), 516-529.

Meirowitz, Adam. 2006. "Designing Institutions to Aggregate Preferences and Information." *Quarterly Journal of Political Science* 1 (4), 373-392.

Ortoleva, Pietro and Eric Snowberg. 2015. "Overconfidence in Political Behavior." *American Economic Review.*

Rabin, Matthew. 1998. "Psychology and Economics." *Journal of Economic Literature* XXXVI (March), 11-46.

Rabin, Matthew and Joel L. Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias." *The Quarterly Journal of Economics* (February), 37-82.

Wason, P. C. 1977. "Self-Contradictions." In P. N. Johnson-Laird and P. C. Wason, eds., *Thinking: Readings in Cognitive Science.* Cambridge University Press, 114-28.

Wason, P. C. 1968. "Reasoning About a Rule." *Quarterly Journal of Experimental Psychology* 20, 273-81.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion.* Cambridge University Press.