C.E.S.S. Working Paper Series

CESS
center for experimental social science

Working Paper #2007-11

Cognition and Strategy: A Deliberation Experiment

Eric S. Dickson, Catherine Hafer, and Dimitri Landa

Key Words:
JEL Classification: C92, D83, C12

19 West Fourth Street, New York, New York 10003, 212-992-8660

The Center for Experimental Social Science

The Center for Experimental Social Science (C.E.S.S.) is an inter-disciplinary research Center at New York University engaged in laboratory experimental work in the social sciences that combines high quality economic theory, psychology and political science. The over arching research tool used in the Center is laboratory experiments in which human subjects engage in decision-making tasks whose purpose is to test both economic theory and/or the properties of proposed or existing economic, political, or social institutions. The working papers in this series reflect these interests.

Cognition and Strategy: A Deliberation Experiment*

Eric S. Dickson† Catherine Hafer‡ Dimitri Landa§

Abstract

A theory of deliberation must provide a plausible account both of individuals' choices to speak or to listen and of how they reinterpret their own views in the aftermath of deliberation. We describe a game-theoretic laboratory experiment in which subjects with diverse interests and information choose to speak or to listen and, after updating their beliefs, vote over a common outcome. An important feature of our strategic setting is that not receiving a specific communication is sometimes just as informative as receiving it. We analyze subjects' deliberative choices and the relationship between these choices, subjects' initial positions and arguments, and individual cognition. Our evidence shows that, although subjects behave instrumentally, their behavior reveals the existence of a cognitive hierarchy defined by differing abilities to grasp the strategic implications of different kinds of information. We trace the consequences of these underlying cognitive differences for individual deliberative choices and for the informativeness of deliberation.

*We are grateful to Becky Morton, Sanford Gordon, and seminar participants at NYU, SUNY Stony Brook, and the Center for Rationality (Hebrew University of Jerusalem) for their helpful comments and suggestions.

†Corresponding author. Assistant Professor, Department of Politics, New York University. Mailing address: 726 Broadway, 7th floor, Department of Politics, New York University, New York, NY 10003. Email: eric.dickson@nyu.edu. Fax: (212) 995-4184.

‡Assistant Professor, Department of Politics, New York University. Mailing address: 726 Broadway, 7th floor, Department of Politics, New York University, New York, NY 10003. Email: catherine.hafer@nyu.edu. Fax: (212) 995-4184.

§Assistant Professor, Department of Politics, New York University. Mailing address: 726 Broadway, 7th floor, Department of Politics, New York University, New York, NY 10003. Email: dimitri.landa@nyu.edu. Fax: (212) 995-4184.

1 Introduction

The free exchange of arguments and the responsiveness of individual and collective decision-making to debate have, in the burgeoning work on deliberative democratic theory, emerged as the new standard of political legitimacy (Cohen 1997). Although the mechanisms supporting this view vary (Elster 1997; Manin 1987; Habermas 1990; and others), most of them suppose that deliberation can have a positive effect by allowing individual and/or collective decision-making to reflect the information or the "best arguments" communicated during interpersonal deliberation, thereby endowing decisions with greater normative acceptability (Cohen 1997; Fearon 1998). That successful, informative, communication will actually take place cannot, however, be taken for granted, and the determinants of such communication are increasingly attracting the attention of game theorists and political economists (Austen-Smith and Feddersen 2006; Gerardi and Yariv 2006; Hafer and Landa 2005, 2007; Meirowitz 2007; Calvert 2006; Patty 2005; etc.). Their key concern in this is an application to the deliberative democratic context of a question that has motivated much of the incomplete information game-theoretic work in political science: when in the course of deliberation do agents find it in their interest to make arguments and to reveal information that is valuable to others?

To the extent that our interest in asking this question lies in explaining intentional individual behavior, it cannot be addressed satisfactorily without also asking what agents do with the information that is made available to them in the underlying strategic process. How do they make sense of it? How do their expectations of how *others* make sense of it affect their perceived incentives to communicate? And what are the ultimate consequences of public deliberation for the quality of decision-making more generally? The nexus of cognition and strategy identified by these questions and the dynamics of deliberation to which it gives rise are the focus of the experiment we describe in this paper. The experiment is based on a simple strategic framework in which individuals have an opportunity to communicate with one another in advance of a collective decision made by simple majority rule. In this framework, individuals may choose either to "speak" or to "listen" as they consider the tradeoffs between influencing others' choices, on the one hand, and gaining potentially useful information on the other. The players' capacities to make sense of the information that would, as a consequence of deliberation, be available to them in the voting stage affect their choices of whether to be senders (speakers) or receivers (listeners). That information is of two kinds: messages that credibly, directly, and fully inform the players of what their preferred policy choices are, and, when those messages are not received, information that could, upon reflection (that is, indirectly), be understood as revealing those choices. A famous literary example illustrates the nature of the inference that is entailed in making sense of the latter kind of information. In Arthur Conan Doyle's "Silver Blaze," Sherlock Holmes points out that the fact that the dog did not bark in the night implies that the crime was not a burglary. Holmes' inference in this and other cases is, as he himself repeatedly announces in Conan Doyle's stories, straightforward. He says, with self-satisfaction and in frustration with the everyman Watson's inability to keep up, "how often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?" (Conan Doyle 1976).

Holmes and Watson differ in their abilities to make an inference from the "null observation." In the context of deliberation, an equivalent of the "null observation" is an 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 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. And of course, he is able to conceive of the latter possibility only if he himself can make inferences from the "null," that is, if he himself is more like Holmes than like Watson.

As this example suggests, individual deliberative choices can be influenced in important ways both by the nature of individual cognition and by the interaction between cognition and strategy. The complexity of the nexus of these two factors in deliberative contexts and the value of isolating the effects of various elements that may be responsible for its consequences point to the benefits of a controlled, socially interactive experiment such as the one we describe.

Our experiment has two basic goals. The first is to learn about the aggregate behavioral patterns or trends present in collective deliberation, which can serve as a guide in positive and normative theorizing. The second, related goal is to explore the microfoundations of judgment formation in deliberation. Different cognitive-behavioral "ideal types," who differ in the way that Holmes and Watson do in how they learn from different kinds of information, will perceive different incentives, which affects not only the way they interpret the messages they receive in the course of deliberation, but also their own strategic deliberative choices. Because different micro-level ways of approaching deliberation have different macro-level behavioral consequences, observed patterns of individual choices regarding speaking and post-deliberative voting shed light on the microfoundations of deliberative practice.

Our analysis sheds light on the deliberative incentives faced by agents in strategically and epistemically different positions— and the way in which agents in these different positions actually behave. For example, consider the divergent incentives faced by ideologically extreme agents and their more moderate counterparts. From an epistemic standpoint, more extreme agents would seem likelier to exhibit a greater sense of certainty about their preferred policy positions than more moderate agents would. In itself, this intuition suggests that more extreme agents might be more inclined towards speaking as opposed to listening than moderates would be, because extreme agents perceive themselves as having less potentially to learn from deliberation (and therefore their "opportunity costs" of speaking are lower). But, as the above example of Jill's interlocutor demonstrates, in some circumstances speaking may tend to push listeners away more effectively than it draws them in.

Our experimental results adjudicate between these conflicting incentives. In accordance with the first expectation, subjects who are more extreme with respect to their pre-deliberative views do in the aggregate tend to speak more than they listen (while more moderate subjects tend to listen much more than they speak). Indeed, subjects with more extreme initial positions exhibit a strong and stable tendency to "overspeak" — that is, to speak even on many occasions when they are more likely to alienate the favorably disposed than to move listeners in their direction. While a small number of our subjects do grasp the dilemma faced by Jill's interlocutor, most do not, when they themselves feel sure of their own positions. In consequence, they both overexpose themselves to the risk that their audiences will move farther away from their preferred positions and, unwittingly, increase the informational value of deliberation.

To account for these and other findings, we turn to the analysis of cognitive-behavioral ideal types and compare individual behavioral profiles to our predictions for the relevant ideal types of agents. This analysis suggests that the observed pattern of communication and voting is best explained by the presence of two types of subjects: (1) those who deviate from the Bayesian ideal by systematically displaying a Watsonian approach to belief updating (to identify the right alternatives, they, like Watson and unlike Holmes - the indisputable Bayesian hero - need to see the direct evidence), and (2) those who may be described as "unreflective Bayesians" - though capable of Bayesian inference themselves, they are unable to appreciate the strategic implications of its possibility. These agents are in a position to take advantage of overspeaking by others while also overspeaking themselves.

2 Relation to Existing Experimental Literature

Our experiment contributes to a growing experimental literature on deliberation. Existing experimental studies exploring different aspects of deliberation can be naturally divided into two categories based on their methodological approaches. Several important studies consider behavior during the course of natural-language, "real issue" deliberation (see, e.g., Fishkin and Luskin 1996; Fishkin, Luskin, and Jowell 2000; French and Laver 2005), involving highly contextual settings. Other studies consider persuasion and choice in more stylized settings; Lupia and McCubbins (1998) is an exemplar of this approach. Both types of research design offer the promise of new insights into deliberation, but both also have distinct strengths and weaknesses which make them more or less suitable for exploring specific research questions.

Because our primary goal is to analyze the relationship between cognition and deliberative behavior, to see how behavior changes as particular aspects of the deliberative environment are varied, and to study the precise factors determining or affecting persuasion, we chose to adopt the second approach, involving a more stylized setting for deliberation. Two key reasons lie behind this choice. First, for the kinds of inferences we wish to make, in the context of a "real issue" deliberation study it would be extremely difficult to measure with adequate precision subjects' prior attitudes, the range of arguments they *would* find persuasive, the specific determinants of persuasion, or the complex and often idiosyncratic cross-issue linkages that may affect individual opinions — or to measure them without prompting change in subjects' views during the process of measurement itself. Second, we wish to study the ways in which a particular individual's deliberative choices change — or don't change — as that individual's strategic position is varied. For both purposes, a stylized environment involving random assignment to different argument and policy positions over a number of periods of deliberation offers access to insights about cognition and the microfoundations of persuasion that would be difficult if not impossible to come by in a more highly contextual experiment.[1]

In this respect the impulse behind our design strategy is similar to that of several experiments reported by Lupia and McCubbins (1998), which study persuasion and choice under a stylized representation of political communication. However, the specific questions we consider and the details of our experimental setup are quite different. For example, we focus on the tradeoff between speaking and listening among members of a deliberative group who must decide whether or not to transmit their partial information about their disparate private interests – and on how subjects learn from this exchange – whereas Lupia and McCubbins focus on the distinction between deceptive and non-deceptive communication between an agent and her principal, and the conditions under which such communication will be taken to be credible.

Our results on the endogenous choices of roles in the deliberative engagement and on the interaction between these choices and individual cognition are, to our knowledge, new to the literature. The same is the case with the finding of overspeaking in a strategic deliberative setting, which has received a further corroboration in a different experimenal setting recently analyzed by McCubbins and Rodriguez (2006). The non-Bayesian account of this finding comports with a large and vigorous experimental literature. A key idea in this literature is the existence of a cognitive bias in favor of one's own currently held convictions (Zaller 1992; Dawes 1998; Rabin 1998; Baron 1994; Taber and Lodge 2006). One potential consequence of such a bias is that only direct and largely unambiguous evidence may be able to overturn one's prior position; 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). Of greatest relevance to us, the literature on the psychology of hypothesis testing (Wason 1968; 1977; Baron 1994, Ch. 13; Taber and Lodge 2006) suggests 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.

A major weakness of the existing experimental literature on cogntive biases as a guide to applied theory in the study of deliberation is that it analyzes such cognitive biases chiefly in the context of individual decision-making - a setting very different from the strategic and interactive environment of political deliberation. As such, it provides at best prima facie clues about the status of these biases in strategic settings - settings which encourage individuals to think through not only their own actions, but their counterparts' likely responses to those actions - and in particular, in strategic settings such as ours, in which, approximating the instances of actual deliberation, subjects receive extensive feedback about the choices made by other players and take on different deliberative roles over the course of a single experimental session. Our study not only demonstrates that such biases can remain intact in such a context, but the novelty of our deliberative setting and experimental design also allows us to pose seldom-asked questions about the microfoundations of choice and learning during deliberation: how do the strategic incentives in deliberation influence individual communication as opposed to voting choices, given the different cognitive tasks that they entail? does the presence of strategically irrelevant types of uncertainty over agents' own preferred policies affect their deliberative choices? are there externalities from deliberative interactions with different cognitive types for agents' own deliberative choices and inferences?

3 The Model of Deliberation

In this section we describe the strategic framework that we employ in our deliberation experiment. The next section develops theoretical predictions on how fully rational individuals might be expected to behave in the different deliberative situations we employ. The remainder of the paper then presents and interprets the results of our experimental investigation.

Individuals who participate in political discourse or in other forms of deliberation and who are approximated by the players in our game can be characterized in terms of the particular set of arguments they may find valid for, or relevant to, the formation of a given judgment. However, like Jill in the introduction, they may not always be aware of or have in mind the full set of arguments that they could or would find to be valid or relevant. When they do not, the communication of arguments that aim to convince "on the merits" - that is, by making apparent the previously missed piece of evidence or a connection between accepted facts or premises[2] - can affect participants' judgments in two ways: (1) by directly bringing to their attention a relevant reason or information that was not an input into their decision-making before, and (2) by (indirectly) making them aware of the arguments or considerations that they find unconvincing and that, therefore, should not be thought to support their best judgments.

These distinctions underscore the value of differentiating between three types of arguments or reasons that play distinct roles in deliberation. Individuals may hold *active* arguments - arguments that they know they find convincing and relevant to the decision at hand. When active arguments provide their holders with only part of the relevant information, they may find themselves being convinced by their *latent* arguments - other persuasive to them relevant arguments that they may hear in the course of the deliberation. (Though the full merits of such arguments become apparent only upon being subjected to them in deliberation, individuals may be aware of their possibility before hearing them.) Finally, active and latent arguments may be distiguished from *foreign* arguments - arguments that are or would, in deliberation, be discovered as unconvincing (if, as (2) above suggests, not necessarily uninformative). In the strategic environment we analyze below, the different epistemic features of active, latent, and foreign arguments are captured by the features of the corresponding similarly designated information *fragments*.

The basic sequence of events is as follows.[3] Individuals who begin in possession of *partial* information about their own best interests are given the opportunity to communicate with one another (the "deliberation stage"); once communication is complete, a vote is held between two potential outcomes, one of which is selected via simple majority rule (the "voting stage"). Each individual receives payoffs that depend on the degree of agreement between her *actual* individual best interests and the election-winning alternative.

Specifically, we consider deliberation within the context of a three-member group. Each member $i \in \{1, 2, 3\}$ has a type $t_i = (t_i^1, t_i^2) \in \{(A, B), (B, C), (C, D)\}$, where $A, B, C, D \in \mathbb{N}$ and $1 \leq A < B < C < D \leq 9$, and an ideal point, or "true number," $x_i^* = 10t_i^1 + t_i^2$, which corresponds to her most preferred outcome. We use the notation $\overline{XY}$ to denote $10X + Y$ - that is, a two-digit number the first digit of which is X, and the second digit of which is Y. The ultimate social outcome is determined by majority rule over a pair of distinct alternatives, $\{y_1, y_2\}$, where $y_1, y_2 \in \{\overline{AB}, \overline{BC}, \overline{CD}\}$ are common knowledge among all group members from the beginning. An individual i's utility from outcome x is linearly decreasing in the "distance" between her true number and the outcome, $u_i(x, x_i^*) = c - |x_i^* - x|$, where c is a constant.

Deliberation has the potential to be persuasive because players do not know their true numbers for certain at the beginning of the game. Instead, each player initially possesses several pieces of information that are relevant to, but which do not necessarily uniquely determine, her true number. First, each player knows that the ideal points of every member of her group (including her own) must be drawn from the commonly known set of possible "true numbers" $\{\overline{AB}, \overline{BC}, \overline{CD}\}$; it may be that all players have different true numbers from within the set, or it may be that two or more players share the same true number. Second, players know the (unconditional) probabilities corresponding to each of these true numbers - how likely it is that a given person would have each of the true numbers if they knew nothing else - all of which are positive. And third, each player knows a "fragment" of her true number, that is, one of the two digits of her true number - for example, "B" or "C" if her true number is $\overline{BC}$ - without knowing whether that fragment is the first or the second digit of her true number. We refer to i's initially known fragment as her "active fragment" or "active argument," $a \in \{t_i^1, t_i^2\}$, and to her initially unknown fragment as her "latent fragment" or "latent argument," $l = \{t_i^1, t_i^2\} \backslash \{a\}$. The set of active fragments known to all group members is common knowledge.

Individuals' first strategic choice is their mode of deliberative participation $\lambda \in \{0, 1\}$, with $\lambda = 0$ capturing the decision to speak and $\lambda = 1$ the decision to listen. A decision to speak entails an attempt to speak to all other members of her group; a decision to listen entails an attempt to listen to those other members of her group who have chosen to speak. Communication is successful only between individuals who have complementary modes of deliberation – e.g., if i speaks and j listens, communication from i to j takes place. Thus, anyone who chooses to speak receives no messages at all, and anyone who chooses to listen sends no messages at all. When complementary speaking and listening choices occur, j (the listener) receives a "message" m_j whose nature depends on both (1) speaker i's active fragment and (2) listener j's true number. If the speaker's active fragment is part of the listener's true number – corresponding either to the listener's active fragment or to her latent fragment – then the "message" listener j receives is simply speaker i's active fragment. However, if the speaker's active fragment is not part of the listener's true number, then the "message" the listener receives is that she has received a "foreign fragment" – that is, she has received a fragment that is not a part of her true number (but that does not explicitly indicate what that fragment is). Thus, given $\lambda_i = 0$ and $\lambda_j = 1$, $m_j = a_i$ if and only if $a_i \in \{t_j^1, t_j^2\}$; otherwise $m_j =$ "foreign."

Modeling agents' deliberative participation λ as a choice between speaking and listening captures in a stark form a common feature of large- and small- scale deliberative processes, in which advocating a policy position by making supporting arguments tends to come at the cost of contemplating the arguments presented by others. The presence of such tradeoffs has both rational and psychological underpinnings. The degree of attachment to a given policy or argument and the extent of the relevant information available to them may be seen to rationally determine agents' willingness to be active information providers. The trade-offs between sending and receiving are also consistent with the variation in the cognitive status of attention and memory that is emphasized in psychological studies of the determinants of persuasion (Zaller 1992; Lupia 2002).

Receiving one's latent fragment in our game is, in effect, equivalent to receiving a direct and immediately convincing argument. Receiving a foreign fragment instead of one's latent argument is equivalent to hearing an unconvincing argument. The analysis of a receiver's response to the observation of a foreign fragment, given the information available to her about the nature of the exchange, and of the senders' anticipation of that response, yields predictions about the interaction between cognition and strategy.

4 Theoretical Predictions

Given the deliberative environment described in the previous section, how can individuals be expected to behave? Because we are interested in the microfoundations of judgment formation in deliberation, it is useful as a baseline to begin by analyzing the behavior of standard Bayesian agents. Before presenting formal behavioral predictions, we illustrate the nature of inference expected from Bayesian agents through the following examples.[4]

Example 1. Suppose it is commonly known that the set of possible true numbers in a three-player group is $\{13, 37, 79\}$ and that the group members' active fragments are 1, 3, and 3. It follows that a given player with the active fragment 3 may have either 13 or 37 as her true number, because both of these contain the fragment 3 (but the true number 79 does not, and so the group must not include a member with that true number). Suppose further that she chooses to listen, and is told that she has received a "foreign fragment." How can such a message prove informative for her? The other player with fragment 3 could not have sent the foreign fragment – if that player had sent her fragment, our listener would have been told that she had received the fragment "3" because "3" is part of her true number. Thus, the "foreign fragment" must have been sent by the subject with the fragment 1; hence the foreign fragment must have been 1; hence "1" must not be part of the listener's true number; hence her true number must be 37. ■

It is clear from this example that, in some circumstances, the receipt of a "foreign fragment" can be as informative as the receipt of the latent fragment. Further, the extent of deduction that is required in order to see this is not particularly demanding. The following two examples illustrate how the possibility of such inferences from the receipt of a "foreign fragment" affects optimal choices regarding speaking vs. listening.

Example 1a. As in the example described above, suppose that it is commonly known that the set of possible true numbers in the group of three players is $\{13, 37, 79\}$ and that the group members' active fragments are 1, 3, and 3, but also that the ultimate vote for a social outcome will be between 13 and 37. Suppose further that there is common knowledge within the group that 37 is more likely than 13; as such, in the absence of further information, the expected utility-maximizing choice for players with the fragment 3 is to vote 37 over 13. Consider first the incentives of the player with fragment 1 in choosing a deliberative strategy. If the players with fragment 3 both speak, the deliberative choice of the fragment 1 player is of no consequence. But that player is weakly better off speaking if at least one of the players with fragment 3 chooses to listen: if the latter receives a message indicating that 1 is her latent fragment, she will know that 13 is her true number, and will now vote for 13 instead of 37 – an improvement for the player with fragment 1 as this now ensures a majority for 13 over 37; if, instead, a listening fragment 3 player receives a message that she has received a foreign fragment, the player with fragment 1 is neither helped nor harmed. That is because a player with fragment 3 will learn that 37, not 13, is her true number *but she would have voted for* 37 *anyway* in the absence of communication because 37 is more likely than 13. Thus, it is a weakly dominant strategy for a player with fragment 1 to speak in this setting. ■

Example 1b. Suppose now exactly the same setting with one exception: it is commonly known within the group that 13 is a priori more likely than 37. The incentives facing the agents are now different. Sending fragment "1" cannot help, and it may hurt, since a Bayesian who receives a foreign fragment will choose 37. If the sender understands that the recipient may make such a deduction, she will strictly prefer not to send in such a situation. ■

We consider two different configurations of active fragments, in which players will face the incentives illustrated in the above examples. In configuration ABB, one player has active fragment "A" while two players have active fragment "B"; in configuration ABC, one player each has active fragment "A," "B," and "C". We will denote the conditional probability that an agent with active fragment B has true number $\overline{AB}$ as $Pr(\overline{AB}|B)$.

In either configuration, players will ultimately vote on $\overline{AB}$ vs. $\overline{BC}$. Players with active fragments A (in the ABB or ABC configuration) or C (in the ABC configuration) can be certain *ex ante* that their most-preferred point in the interval $[\overline{AB}, \overline{BC}]$ will be one of the endpoints of that interval ($\overline{AB}$ when A and $\overline{BC}$ when C), whereas players with active fragment B will have a true number whose *expected* value lies in the interior of that interval. Given the implicit left-right "connectedness" of the fragment/argument profile (there is no fragment that is shared by the types associated with the "ends" of the $[\overline{AB}, \overline{BC}]$ policy interval that is not also shared by the types associated with its interior), and the intuitive ambiguity-based distinction between ideological moderateness and extremeness suggested in the introduction, we will refer, in this context, to the A and C fragments as "extreme" and to the B fragment as "moderate," while referring to the players possessing the relevant active fragments as "extremists" and "moderates," respectively.

Our solution concept is weak dominance, which here generates unique predictions with respect to both deliberation and voting. Table 1a summarizes the unique equilibrium predictions for speaking and listening by players in various deliberative settings, while Table 1b summarizes the unique equilibrium predictions for voting choices in these same deliberative settings, given the different possible profiles of messages received during deliberation.[5]

The logic of the informal examples induces broad patterns in the tables describing optimal behavior. When, as in Example 1a, Bayesian agents would perceive an incentive to speak, we will refer to this as being *the Speaking Case* for the relevant A or C agent (in either configuration). And when, as in Example 1b, such agents would perceive an incentive to listen (so as not to alienate other favorably disposed agents), we will refer to this as being *the Listening Case* for the relevant A or C agent (in either configuration).

5 Experimental Sessions

The experiment was carried out at the NYU Center for Experimental Social Science (CESS). Our results come from data collected in two experimental sessions involving 18 subjects each, for a total of 36 subjects. Subjects signed up for the experiment via a web-based recruitment system that draws from a broad pool of potential participants; individuals in the subject pool are mostly undergraduates from around the university, though a smaller number came from the broader community. We did not recruit from our classes, and all subjects gave informed consent according to standard human subjects protocols. Subjects interacted anonymously via networked computers; the experiment was programmed and conducted with the software z-Tree (Fischbacher 1999).

Each of our experimental sessions consisted of 30 rounds, where each round corresponds to a single play of our deliberation-and-voting game. Sessions lasted approximately 90 minutes, and on average subjects earned US$26.56, including a showup fee of US$7. At the beginning of each session, a hard copy of the experimental instructions (which can be found in the reviewers' supplemental appendix) was distributed to each of the subjects; the instructions were also read aloud in an attempt to induce common knowledge. Before beginning the experiment, subjects took an on-screen quiz consisting of six questions in order to test their understanding and as a further means of inducing common knowledge of the deliberation framework; subjects were given immediate feedback as to the correct answers. After all 30 rounds were complete, subjects then completed a post-experiment debriefing survey.

At the beginning of each round, subjects were randomly rematched into a new group of 3 members and were assigned a new role in a new deliberative environment. Roles were assigned in a quasi-random fashion designed to ensure wide experience of different strategic incentives. Because of our interest in the nexus of cognition and strategy in deliberation, and our desire to ensure that our inferences about subject behavior and cognition are not merely an artifact of inexperience in our deliberative setting, we devised an *ex ante* division of the session into two parts for the purposes of analysis. Subjects were to be deemed "inexperienced" during periods 1-12; the quasi-random assignment to roles and situations ensured that during this initial phase each subject was exposed three times each to Speaking Case (ABB), Listening Case (ABB), Speaking Case (ABC), and Listening Case (ABC), and played every role within each of these situations. This was done in order to ensure that, for each subject, many periods took place following this diverse exposure to every permutation of roles and incentives, so that subjects' behavior when "experienced" could be measured. In accordance with our initial plan, most of our analyses thus concern the "experienced" periods 13-30, to strengthen our confidence that any inferences about behavior or cognition are not merely an artifact of the novelty of the scenario during the initial 12 periods of learning. This division of the session into two halves is for analysis purposes only; subjects were presented with the same interface, performed the same tasks, and received payoffs according to the same formula in all of the periods. Overall, each subject participated in the following distribution of situations: 7 Speaking Case (ABB); 7 Listening Case (ABB); 8 Speaking Case (ABC); 8 Listening Case (ABC). The specific parameters of each situation (the set of possible true numbers, the unconditional probabilities, etc.) were different from round to round, but all groups in a given round were set in the same deliberative environment.[6] Subjects' payoffs were in cents, with $c = 80$, so that subjects received 80 cents in each round *minus* 1 cent for each "unit of distance" between the winning number and their own true number.

After taking their last action in each round (voting), subjects were given as feedback the winner of the vote as well as the vote totals, what their actual true number had been in that round, and their payoff for the round in cents. (This feedback gave subjects who chose to listen an opportunity to learn about the correctness of their vote (and therefore about how well they were using information), and it gave subjects who chose to speak an opportunity to learn something about the behavior of their intended audience.) Subjects who spoke received no feedback about others.

6 Experimental Results: Aggregate Level

6.1 Deliberative Choices: Speaking and Listening

As the discussion in the theoretical results section demonstrates, individuals in our deliberation situation face differing strategic incentives depending on the active fragment that is known to them; the distribution of active fragments across other group members; and the relative likelihoods of different true numbers. Table 2a contains relevant data from our experimental sessions compiled in the same format as Table 1a.

The first thing to note is a systematic difference in deliberative behavior between agents depending on the nature of their active fragment – in particular, whether or not an individual's active fragment is more "moderate" or more "extreme." In the settings we describe, listening is always a weakly dominant strategy for individuals with moderate active fragments, while either speaking or listening can be weakly dominant for individuals with extreme active fragments, depending on the situation and their cognitive approach to deliberation.[7] Our first conclusion indicates that the distinction between the deliberative behavior theoretically expected of moderate and extreme agents is strongly apparent in our data.

Conclusion 1. Extremists Speak More, Moderates Listen More. *Subjects with more* *extreme active fragments speak more frequently than they listen, while subjects with more moderate active fragments listen more frequently than they speak.*

The difference in deliberative behavior between "moderates" and "extremists" is striking, as can be seen in Table 2a. Over the last 18 periods of the experiment, once subjects had had some experience of each kind of deliberative role and setting, subjects with a moderate ('B') active fragment chose their weakly dominant strategy – to listen – more than 96% of the time in each of the deliberative situations. Pooling across deliberative situations, they did so fully 98.1% of the time (306/312).[8] In contrast, subjects with an extreme ('A' or 'C') active fragment, whose incentives differed across different cases, chose to *speak* between 54.2% and 79.2% of the time in different deliberative situations over the last 18 periods of the experiment. This strong difference in the tendency of moderates and extremists to adopt different modes of deliberation is not only of substantive interest, but also provides one clear piece of evidence that subjects' deliberative behavior was responsive to the details of their situations.

Our second conclusion reports the way in which their behavior varies between the Speaking Case and the Listening Case – that is, between situations in which a Bayesian agent would perceive that speaking as opposed to listening is her weakly dominant strategy.

Conclusion 2. *At the aggregate level, subjects with an extreme fragment chose to speak more often in the Speaking Case than they did in the Listening Case. However, they chose to speak substantially more than half the time in both of the Cases, suggesting that population-level average speaking incidence substantially exceeds the predicted incidence for the Bayesian agents. Learning over the course of the experimental sessions significantly increases the incidence of speaking in the Speaking Case but does not significantly alter the incidence of speaking in the Listening Case.*

Table 2a indicates that, in the last 18 periods, our subjects chose to speak between 54.2% and 66.7% of the time in each of the three settings corresponding to the Listening Case, for an overall average of 60.1% (101/168), in sharp contrast to the Bayesian prediction of 0%. Subjects spoke more often, between 73.3% and 79.2% of the time, in the three other settings corresponding to the Speaking Case, for an overall average of 76.8% (129/168), as against the predicted 100% for Bayesian agents. The null hypothesis that there is no difference in behavior between the Speaking Case and the Listening Case in these last 18 periods can be rejected decisively ($Z = 3.287$, $p < 0.001$). Thus, while predicted deliberative behavior for moderates conforms closely to the theoretical expectations for Bayesians (as seen above, 98.1% observed versus 100% predicted listening), systematic deviations from the Bayesian predictions are observed for extremists, and by far the strongest deviations take the form of "overspeaking" by extremists in the Listening Case.

The importance and robustness of this "overspeaking" in the Listening Case is underscored by considering time trends in subjects' deliberative behavior. As shown in Table 2b, during the first 12 periods of the experiment, subjects in the Speaking Case chose to speak 62.0% of the time (67/108), whereas in the Listening Case they did so 63.0% of the time (68/108) – that is, there is no significant difference in deliberative behavior across Cases when subjects are inexperienced, but as shown above there is a substantial and significant difference once they are experienced. Further, it is important to note that the difference between the first 12 and the last 18 periods is due almost entirely to a shift in subjects' behavior in the Speaking Case. In the Speaking Case, the null hypothesis that behavior does not change from the first 12 periods (62.0%) to the last 18 periods (76.8%) is rejected ($Z = 2.636$, $p < 0.01$), whereas in the Listening Case, the null that behavior does not change from the first 12 periods (63.0%) to the last 18 (60.1%) cannot be rejected ($Z = 0.473$, $p = 0.64$). This observation strengthens confidence that subjects' failure to behave according to Bayesian predictions – and in particular, to exhibit a pronounced pattern of "overspeaking" in the Listening Case – is not entirely a result of inexperienced misunderstanding, because the behavior persists over the course of an experiment with a large number of rounds.

As noted in our theoretical predictions, we employ deliberative settings with different distributions of active fragments; this element of variation in our design gives us further ability to probe the extent to which subjects' *strategic* incentives influence their behavior, as opposed to other factors that are strategically irrelevant in our framework. A striking feature of our experimental results is the degree of instrumentalism that subjects' choices may be seen to reveal, despite their apparent suboptimality:

Conclusion 3. Subjects With Identical Strategic Incentives Deliberate Similarly Across Informationally Distinct Deliberative Contexts. *'B' agents, who have a dominant strategy to listen, do so at statistically identical rates in ABB and ABC situations. Further, 'A' agents and 'C' agents who face strategically identical problems within the ABC situation also differ little in their deliberative behavior, even though 'A' agents can be ex ante certain of their true numbers while 'C' agents cannot.*

Thus, subjects' strategic incentives matter more to their behavior than do such descriptive aspects of the deliberative problem as the degree of *ex ante* certainty they have about their true numbers. Even when listening is crucial to the subjects' ability to learn arguments necessary for determining their ideal policies (their unconstrained optima), their decisions regarding speaking and listening tend to be induced by the particular voting agendas - that is, by the extent to which what they could learn in deliberation is *necessary* for determining their "induced" preferences over the alternatives on the agenda (their constrained optima).

The ABB and ABC situations differ in their degrees of descriptive and inferential complexity – a 'B' agent may have more difficulty discerning the meaning of a foreign fragment under ABC than under ABB, for example, while a 'C' agent who has a strategic incentive to send may find this more difficult to understand than a comparable 'A' agent, because the 'C' agent does not know her true number for certain while the 'A' agent does. Table 2a indicates that for experienced subjects in the last 18 periods, most of the variation in deliberative behavior across circumstances is captured by the strategic logic of the model rather than by contextual factors such as these that are strategically irrelevant. 'B' agents listen with remarkable consistency across deliberative settings – the variation between 96.7% and 99.0% is statistically insignificant. Similarly, 'A' and 'C' agents who, as Bayesian agents, would perceive an incentive to speak, also are remarkably consistent across contexts, doing so between 73.3% and 79.2% of the time, another insignificant difference. Finally, 'A' and 'C' agents who, as Bayesian agents, would perceive an incentive to listen, appear to do so a bit less consistently across contexts, varying from 54.2% to 66.7% of the time. In particular, such 'C' agents under ABC speak a bit less than do such 'A' agents under ABC, which may reflect 'C'

agents' relative lack of true knowledge about their actual true number. However, even this difference is statistically insignificant.

In order to further probe the extent to which behavior may have been affected by strategically superfluous descriptive features of the deliberative environment, we carried out a probit regression analysis of deliberative choice for those subjects with an extreme active fragment. Results from the regression, which was carried out for periods 13-30, are contained in Table 3. The dependent variable is the dichotomous choice between listening (1) and speaking (0). The independent variables depict a variety of features of the deliberative environment as well as a time trend variable. The regression results indicate that only subjects' strategic incentives – as categorized by the distinction between the Speaking and the Listening Cases – significantly affect behavior. Other factors, such as the distinction between ABB and ABC situations; the distinction between "left-" and "right-handed" situations (e.g., ABB vs. CCD); and strategically irrelevant features of the true numbers' and probabilities' specific values all have a statistically insignificant effect on behavior. The time trend variable is also insignificant, indicating no strong learning trend after period 13. These results give us confidence that subjects' choices were affected by their strategic incentives but were not unduly influenced by other factors. With these results in mind, for the remainder of the section we separately pool all decisions together that are within a common case (Speaking or Listening), and restrict our attention to periods 13-30.

6.2 Voting Choices

We begin our consideration of the voting data by noting the frequencies with which subjects who play a particular role in a given round of the experiment are exposed to a variety of different deliberative outcomes. These frequencies are of course determined jointly by chance and by the deliberative strategies chosen by a given subject and his or her counterparts. For a given individual who has been assigned the 'B'-type fragment, three separate results of deliberation are about equally likely to be observed: the receipt of no fragment (29.9% of the time); the receipt of a foreign fragment only (26.7% of the time); and the receipt of the subject's latent argument only (25.0% of the time). The conjunction of the latent argument and a foreign fragment was considerably less frequent (15.7% of the time), and the rest of the possible outcomes essentially not present (from at or below 1% of the time).

The frequencies of the deliberative outcomes perceived by individuals who had been assigned the 'A'- or 'C'-type fragments reflect the different deliberative choices made by these agents compared to those who possess the 'B' fragment. Most of the time (84.2%) 'A' or 'C' agents receive no signal at all – naturally the case, given that 'B' agents almost always listen and the 'A' and 'C' agents themselves speak more often than they listen. 'A' and 'C' agents receive only a foreign fragment 13.6% of the time and observed other outcomes very rarely.

Along with our theoretical expectations, these distributions of deliberative outcomes lead us to expect that we will learn the most about subjects' behavioral types from voting behavior by observing the choices made by the 'B' agents. Our next two conclusions focus on those choices. The first conclusion checks how well the subjects grasp the problem that is posed to them and on their ability to make the elementary probabilistic inferences they are expected to perform:

Conclusion 4. Subjects Almost Always Use Dominant Voting Strategies When Receiving No Signal or When Receiving Their Latent Fragment. *At the aggregate level, when subjects with a 'B' (moderate) active fragment receive no signal, they almost always vote according to their efficient prior belief, indicating a strong understanding of the probabilities' actual meanings. Further, when subjects with a 'B' active fragment receive their latent fragment, they almost always vote correctly (for their actual true number).*

The data supporting this Conclusion can be found in Table 2c. Aggregating across deliberative settings, subjects with a 'B' (moderate) active fragment who receive no signal at all vote in accordance with their prior belief 94.9% of the time (150/158). Further, subjects with a 'B' active fragment who receive their latent fragment vote correctly (for their now-known true number) 95.0% of the time (209/220). Taken together, these statistics indicate a striking degree of understanding among our subjects both of the meaning of the unconditional probabilities they were given and of the meaning of latent fragments. The significance of this conclusion is largely instrumental: it puts in perspective our next conclusion, which describes voting behavior by those 'B' agents who receive only a foreign fragment:

Conclusion 5. Subjects Often Fail to Learn From Informative Foreign Fragments. *At the aggregate level, when subjects with a 'B' active fragment receive only a foreign fragment in the Listening Case (ABB), the pattern of voting behavior indicates a strong and systematic failure to infer that their true numbers are not* $\overline{AB}$. *In the cases in which receiving a foreign fragment should indicate to subjects that they should vote according to their priors, they vote correctly almost all of the time. But in the cases in which receiving a foreign fragment should indicate to subjects that they should vote against their priors, they vote correctly much less often (slightly more than half of the time).*

This conclusion exploits the differences in the information content of foreign fragments in different deliberative settings. In the Speaking Case (ABB), the 'B' agent's prior indicates that, in the absence of further information, she should prefer to vote for $\overline{BC}$ over $\overline{AB}$. If, in this setting, a Bayesian 'B' agent receives a foreign fragment, she will understand that the foreign fragment must have been sent by the 'A' agent; that her own true number cannot contain the 'A' fragment; and therefore that her own true number must be $\overline{BC}$. The same is also true in an ABC situation: a Bayesian agent who understands the strategic incentives faced by 'A' and 'C' types will expect that, if one foreign fragment only is received, that it would have most likely been sent *by the agent who was trying to change her mind* – and therefore should not affect her ultimate vote.

However, the situation in the Listening Case (ABB) is quite different. The 'B' agent's prior belief inclines her to vote for $\overline{AB}$ over $\overline{BC}$. However, unlike in the cases discussed above, receiving a foreign fragment puts the 'B' agent's prior beliefs in tension with the information represented by it. She now ought to vote against her prior belief, because she would know with certainty that her true number could not contain fragment 'A.' As such, a comparison of the voting behavior of 'B' agents who have received only a foreign fragment between the Listening Case (ABB) and the other cases provides a direct test for the hypothesis that agents fail to make inferences from unpersuasive but informative arguments. The results of this comparison can also be found in Table 2c.

The rate with which subjects fail to vote "correctly" upon receiving a foreign fragment that indicates against their prior is striking. Taken over all 30 periods, subjects vote correctly against their prior 56.2% of the time (18/32) in the Listening Case (ABB), but they vote correctly with their prior 91.7% of the time (100/109) in the other cases. The null hypothesis that these success rates are identical is soundly rejected by a difference-of-proportions test ($Z = 4.778$, $p < 0.0001$). Restricting attention to the last 18 periods (13-30), subjects vote correctly (against their prior) 62.5% of the time (10/16) in the Listening Case (ABB), but they vote correctly (with their prior) 93.4% of the time (71/76) in the other cases. In this instance the null hypothesis that these success rates are identical is also overwhelmingly rejected ($Z = 3.465$, $p < 0.001$).

At the risk of repeating ourselves, it is worth emphasizing that the high rate of errors in voting that we note above takes place in the context of the Listening Case (ABB). If a 'B' subject receives a foreign fragment it is *only* the 'A' agent from whom the foreign fragment could have come. While the information contained in such a foreign fragment is not as transparently *labeled* as the other sorts of unambiguous information – the receipt of a latent fragment, or being endowed with an "endpoint" fragment – that lead to correct voting rates in excess of 95%, it is in actuality no less informative. The ability of subjects to make correct inferences where they do, but *not* to make them when receiving such foreign fragments, fits with the logic of a violation of *Negative Introspection* (the ability to learn from what does *not* happen or what one does *not* know—see, e.g., Binmore 1990, pp. 108-110).

7 Re-examining Microfoundations

7.1 Cognitive Types and Individual Behavioral Profiles

The five conclusions presented above all concern aggregate-level data that pools together the behavior of all of our experimental subjects. Of course, aggregate-level findings about deliberative choices may result from a wide variety of different individual-level behaviors. Among other things, the given aggregate-level findings can be generated either by a homogeneous population of individuals who behave probabilistically in the same way that population averages do, or by various configurations of highly heterogeneous individual behavior. In this section we address our experimental findings at the individual level. Although the conclusions reported in the previous section indicate substantial presence of suboptimal behavior, they also show that the behavior is far from random, naturally raising the question of what may account for it.

As we noted in the introduction, our explanation hinges on individuals' understandings of the epistemic and strategic implications of indirect evidence, variation in which gives rise to distinct cognitive types in the population of subjects. Before examining the individual-level data in detail, we first consider and reject another account that may, at first blush, seem consistent with the evidence presented in the previous section. This account turns on the presence of a population of agents who have so little understanding of the problem that the unconditional probabilities of the true numbers have no meaning for them at all. Such agents might not vote according to the relative values of the unconditional probabilities – that is, they may ignore the distinction between the Speaking and the Listening Cases, perhaps voting merely randomly. However, even such agents could probably be expected to understand their actual true numbers if they possessed both the active and the latent fragments. If present in the population in sufficient numbers, such agents might induce a Bayesian subject to speak – regardless of the Case, and so potentially consistent with our finding of considerable "over-speaking" in the Listening Case – if she believed that the population of such deviant agents was so large as to make the advantages of persuading them outweigh the potential risks of dissuading fellow Bayesians.

This explanation is implausible for several reasons. First, subjects almost universally gave correct answers to the quiz question about the unconditional probabilities (34/36, 94.4%).[9] Second, as noted in Conclusion 4, subjects with the moderate fragment who receive no signal at all practically always vote according to their (unconditional) prior belief. These two points suggest that very few if any agents actually possess the particular deficiency in understanding, described above, that would induce Bayesian agents always to speak. Finally, subjects' responses to our post-experiment survey provide further evidence that this view was, in fact, widely shared by them.[10]

The implausibility of the explanation that relies on agents' beliefs that their counterparts are "too irrational" points to the appeal of an explanatory strategy that we pursue in the remainder of this section. This strategy is to identify analytically plausible cognitive-behavioral ideal types and compare our predictions for their respective behaviors with the complete individual behavioral profiles of our experimental subjects. A natural point of departure in constructing such ideal types is the Bayesian agents that appear in standard game-theoretic models and that underlie our initial theoretical predictions. Such agents, who update their beliefs efficiently using Bayes' Rule and who, like Holmes, understand the logical implications of any information they possess and incorporate them into their beliefs, may be thought of as being at the top step of the hierarchy of different types that possess different levels of understanding of the strategic environment and use the information they receive with different levels of insight.[11]

Agents below the top step in this hierarchy differ from the Bayesian agents in that they lack some degree of logical omniscience: they are unable to recognize some of the logical implications of certain pieces of information they encounter, and, because of this, make suboptimal choices. In the context of the analytical tasks faced by our subjects, there is a natural line that can be drawn between those agents who are capable of making a Bayesian inference from a foreign fragment and those who are not. With this line in mind, we can define a cognitive type - we refer to it as *Watsonian* - that differs from the Bayesian ideal type in that agents that belong to it are not negatively introspective, i.e., they do not make the necessary inferences in order to correctly update their beliefs upon receiving a foreign fragment, even though (as we have seen) foreign fragments can be informative, and like Holmes' Dr. Watson, this type of agents have a full command of of all other aspects of the deliberative environment.

The interactive nature of the decision setting in our experiment suggests the possibility of a further cognitive type that falls between Watsonians and Bayesians in the cognitive hierarchy. The intuition for this possibility is that the correct identification of the optimal speaking behavior for the Bayesian agents effectively presupposes the correct solution of a somewhat easier problem of inference as a listener. Thus, there may exist a cognitive-behavioral ideal type that is capable of a negatively introspective inference herself, but that is different from the standard Bayesian type in being unable to grasp the implications of that possibility for the strategic choice of speaking vs. listening. We refer to the agents who belong to this ideal type as *Unreflective Bayesians*.

Bayesians, Unreflective Bayesians, and Watsonians do not, course, exhaust the set of possible ideal types. We can also imagine agents who are below Watsonians in the cognitive hierarchy: for example, agents who behave randomly, or who have an idiosyncratic and more gravely mistaken understanding of the deliberative environment than Watsonian agents do. Because we do not possess strong intuitions for which of the many such possible idiosyncratic behaviors are most likely to be observed in actual subject populations, we do not provide a specific characterization of the behavior we would expect of such agents and instead refer to them collectively as *Deviant*.

Because neither Watsonian nor Unreflective Bayesian agents are able to conceive of the possibility of their fellow group members using “against them” the information they might provide by choosing to send, they are indifferent between sending and receiving in all circumstances in which they are initially endowed with an extreme fragment and the moderate agent’s pre-communicative beliefs are in their favor. In Table 1a, which describes the equilibrium speaking and listening choices of Bayesian agents, these circumstances correspond to the choices marked in boldface. In fact, these choices are the only differences in optimal speaking sub-strategies between them and the Bayesian agents. (The formal derivation of the Watsonians’ optimal strategies is in the reviewers’ supplemental apendix; the Unreflective Bayesians’ speaking sub-strategies are the same and are justified by the same argument). The only difference between Watsonians and Unreflective Bayesians concerns the voting choice in the Listening Case (ABB) when receiving a foreign fragment - i.e., the situation that corresponds to the bold font entry in Table 1b. Whereas the Bayesians and the Unreflective Bayesians would switch from the prior optimum of $\overline{AB}$ to $\overline{BC}$, Watsonians would continue preferring $\overline{AB}$.

Given these predictions, we can now look for evidence for each of these ideal types in the individual-level data. Keeping in mind the observational equivalence between Unreflective Bayesians’ and Watsonians’ deliberative choices (i.e., speaking vs. listening decisions), we can state the following conclusion:

Conclusion 6. Inferences about Behavioral Types: Individual Deliberation Data. *At the individual level, there is substantial heterogeneity in the deliberative behaviors demonstrated by subjects when cast in the role of extremist. Roughly half of the subjects demonstrate behavior that is consistent or nearly consistent with the behavior theoretically associated with the Watsonian and the Unreflective Bayesian behavioral types. A much smaller number of subjects demonstrate behavior that is consistent or nearly consistent with the behavior theoretically associated with the Bayesian behavioral type. The remaining subjects display highly idiosyncratic behaviors, some of which involve the repeated selection of dominated strategies.*

Table 4 contains the individual-level deliberation choices in rounds 13-30 for subjects who possess extreme active fragments. (Note: subjects 1-18 took part in Session 1; subjects 19-36 took part in Session 2.) We restrict our attention to rounds 13-30 in an attempt to obtain a relatively meaningful, if rough, classification of behavior that takes into account an initial period of learning by subjects.[12] In order to organize this data, we employ the following system of categorization. If a subject behaves in precisely the same way as a given behavioral type would, or deviates from this behavior by no more than one choice in the Speaking Case and by no more than one choice in the Listening Case, s/he is described as being of that behavioral type. For the Bayesian type, this involves speaking in all or all but one of the Speaking Case circumstances, and listening in all or all but one of the Listening Case circumstances. For the Watsonian and the Unreflective Bayesian types, this involves speaking in all or all but one of the Speaking Case circumstances, and speaking in at least half of the Listening Case circumstances.[13] Behavior that is not classified according to either standard will be discussed later.

This system of categorization yields the following results. Of the 36 subjects, 19 can be uniquely categorized as belonging to the Watsonian/Unreflective Bayesian types (15 of these 19 make choices that correspond exactly to possible choice behavior for a Watsonian and the Unreflective Bayesian agents as we have defined them). Only 3 subjects can be uniquely categorized as belonging to the Bayesian type (each of them spoks in every Speaking Case situation but listened in every Listening Case situation). These unambiguous assignments of subjects to the Watsonian/Unreflective

Bayesian or to the Bayesian type account for 22 of the 36 subjects.

Of the remaining 14 subjects, 2 subjects "fall on the boundary" between Bayesian and Watsonian/Unreflective Bayesian types; because these subjects were exposed to a small number of instances of the Listening Case, they were both simultaneously one choice away from the Bayesian classification and one choice away from the Watsonian/Unreflective Bayesian classification. As such, these subjects are left uncategorized. Two other subjects fell one further choice away from one of the behavioral types than the categorization system allowed; one of these was "leaning" towards a Bayesian classification and the other was "leaning" towards a Watsonian/Unreflective Bayesian classification. The remaining 10 subjects, who in Table 4 are labelled "deviant," did not fit the pattern of either behavioral type, engaging in especially perverse behavior – for example, always listening, regardless of the case – or exhibiting a pattern of choices that appears essentially random.

We next consider how the hierarchy of ideal types we proposed above comports with subjects' voting behavior. Because virtually all subjects vote consistently with their prior in the absence of a message informing them of having received a foreign fragment, the relevant variation concerns the circumstances in which they do receive such a message. Table 4 also reports individual-level information regarding subjects' voting choices. Our analysis of the individual-level voting data gives rise to the following conclusion:

Conclusion 7. Consistency of Behavioral Types Between Deliberation and Voting. *At the individual level, when correlating deliberative behavior with voting behavior, subjects who behave as Bayesian agents in choosing a deliberative strategy vote correctly upon receiving only a foreign fragment. When pooled together, subjects who behave as Watsonian/Unreflective Bayesian agents in choosing a deliberative strategy vote correctly less often but more often than the subjects classified as Deviant.*

If the behavioral types we have defined are a reasonably good description of how subjects view deliberation and form judgments about their interests, and they correctly instantiate a true latent cognitive-behavioral hierarchy present in the population of subjects, it should be the case that the deliberative choices made in the first stage exhibit consistency with the voting choices made in the second stage. In order to test for this linkage, we compared subjects' individual-level vote choices for the 32 instances of the Listening Case (ABB) contained in Table 2c with their behavioral classifications – derived from deliberative choices made in different rounds – contained in Table 4. The result was as follows: agents classified as Bayesian voted correctly 100% of the time (4/4); agents classified as Watsonian and Unreflective Bayesian voted correctly 57.9% of the time (11/19); and agents classified as Deviant voted correctly 28.6% of the time (2/7). As such, the pattern of subjects' ability to use information from foreign fragments to vote correctly is consistent with the behavioral classifications based on deliberative choice, as one would expect if stable behavioral types were an accurate model of subjects.[14]

It is worth emphasizing the suggestive nature of the pattern of voting behavior pattern exhibited by subjects classified as Watsonian or Unreflective Bayesian on the basis of their deliberative choices. Subjects seldom confront conditions that allow individual classification based on voting behavior, but aggregate voting patterns lend themselves to the interpretation that Watsonian and Unreflective Bayesian types are likely both present in significant numbers. Indeed, subjects vote correctly in 57.9% (11/19) of relevant circumstances; if subjects were homogeneously Watsonian (vs. Unreflective Bayesian) this figure would be 0% (vs. 100%). As we have seen, subjects nearly always vote correctly outside of Listening Case (ABB); the single, sharp downward deviation in voting accuracy in exactly the place it would be expected of Watsonians does seem to indicate their presence in the sample. However, while 57.9% is far from 100%, it is also far from 0%, pointing to the likely presence of Unreflective Bayesians as well.[15]

8 Conclusion

Constructing empirically relevant theories of deliberation requires that we take seriously how the content of and response to the voiced arguments may be affected by individuals' perceptions of each other in the context of the underlying political issue or problem. What may be expected from this relationship is a function of the particular features of what we described in the introduction as the "nexus of cognition and strategy." As an attempt to shed light on those features, the experiment described in this paper offers an account of deliberation that suggests elements of consistency with, but also fundamental and systematic differences from, equilibrium predictions for the underlying strategic model. Although we find that subjects are quite instrumental in their choices—they pay attention to payoff-relevant and ignore payoff-irrelevant features of the deliberative environment—they also exhibit a persistent tendency to ignore the strategic implications of their speech in choosing a mode of deliberative participation.

By far the strongest manifestation of this tendency is subjects' engagement in public argumentation when their optimal choices counsel silence. Two consequences of this outcome for deliberative practice deserve particular mention. The first consequence is behavioral, and it receives strong, direct support in the experiment: more ideologically extreme agents should, *as a rule*, be expected to speak rather than listen, while more ideologically moderate agents exhibit the opposite behavior. The second consequence concerns the informational content of deliberation. While "overspeaking" connotes suboptimality of individual deliberative choices, it also indicates that deliberative engagement may lead to greater levels of socially available information for post-deliberative decision-making than would exist under optimal speaking and listening choices. Individual suboptimality here has a social silver lining, and this conclusion can be seen to bolster the prospects of deliberation-based accounts of democracy. Still, our results also show that while some agents may take advantage of this "unexpected" informational windfall, others – in our experiment, almost a half of the participants – may not.

When combined with the expectation of greater speaking by those with the strongest prior ideological bias, this kind of inefficient learning has an important implication for understanding the phenomenon of post-deliberative group polarization (Mendelberg 2002, Sunstein 2002). Watsonian failures to learn from "the null event" mean that biased speech in closed biased groups is more likely to move listeners towards the speaker than away from her.

Our analysis of individual behavioral profiles provides strong evidence that the degree of responsiveness to "null events" is the critical systematic factor explaining variation in subject behavior.

But while many subjects display quintessentially Watsonian epistemic unresponsiveness to such events, we also find evidence of a fundamentally strategic type of cognitive shortcoming: agents who are epistemically *responsive* to null events, but who fail to recognize the strategic implications of such responsiveness on the part of others, and who, like Watsonians, "overspeak."

References

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

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

[3] Binmore, Ken. 1990. *Essays on the Foundations of Game Theory.* London: Basil Blackwell.

[4] Calvert, Randall L. 2006. "Deliberation as Coordination Through Cheap-Talk." Washington University Mimeo.

[5] Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics* 119(3), 861-898.

[6] Cohen, Joshua. 1997. "Deliberation and Democratic Legitimacy." In J. Bohman and W. Rehg, eds., *Deliberative Democracy: Essays on Reason and Politics.* MIT Press, 67-92.

[7] Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica* 69(5), 1193-1235.

[8] Doyle, Arthur Conan. 1976. *The Complete Sherlock Holmes.* Doubleday.

[9] 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.

[10] Elster, Jon. 1997. "The Market and the Forum: Three Varieties of Political Theory." In J. Bohman and W. Rehg, eds., *Deliberative Democracy.* Cambridge: MIT Press, 3-34.

[11] Fearon, James. 1998. "Deliberation as Discussion." In Jon Elster, ed., *Deliberative Democracy.* Cambridge University Press.

[12] Fischbacher, Urs. 1999. "z-Tree - Zurich Toolbox for Readymade Economic Experiments - Experimenter's Manual." Working Paper Nr. 21, Institute for Empirical Research in Economics, University of Zurich.

[13] Fishkin, James, and Robert C. Luskin. 1996. "The Deliberative Poll: A Reply to our Critics," *Public Perspective* 7(1), 45-49.

[14] Fishkin, James, Robert C. Luskin, and Roger Jowell. 2000. "Deliberative Polling and Public Consultation," *Parliamentary Affairs* 53, 657-666.

[15] French, Damien, and Michael Laver. 2005. "Participation Bias and Framing Effects in Citizens' Juries." Paper, Annual Meeting of the American Political Science Association.

[16] Gerardi, Dino and Leeat Yariv. 2006. "Deliberative Voting." *Journal of Economic Theory*, forthcoming.

[17] Glazer, Jacob and Ariel Rubinstein. 2005. "On the Pragmatics of Persuasion: a Game Theoretical Approach." Tel-Aviv University Mimeo.

[18] Habermas, Jurgen. 1990. *Moral Consciousness and Communicative Action.* Cambridge: MIT Press.

[19] Hafer, Catherine and Dimitri Landa. 2005. "Deliberation, Ideological Bias, and Group Choice." New York University Mimeo.

[20] Hafer, Catherine and Dimitri Landa. 2007. "Deliberation as Self-Discovery and Institutions for Political Speech." *Journal of Theoretical Politics*, forthcoming.

[21] Lipman, Bart and Duane J. Seppi. 1995. Robust Inference in Communication Games with Partial Proveability. *Journal of Economic Theory* 66, 370-405.

[22] 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.

[23] Lupia, Arthur. 2002. "Deliberation Disconnected: What It Takes to Improve Civic Competence." *Law and Contemporary Problems* 65 (3), 133-50.

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

[25] McCubbins, Mathew D. and Daniel B. Rodriguez, 2006. "When Does Deliberating Improve Decision Making?" UCSD Mimeo.

[26] Manin, Bernard. 1987. "On Legitimacy and Political Deliberation." *Political Theory* 15 (3), 338-68.

[27] Meirowitz, Adam. 2007. "In Defense of Exclusionary Deliberation: Communication and Voting with Private Beliefs and Values." *Journal of Theoretical Politics*, forthcoming.

[28] Patty, John. 2005. "Arguments-Based Collective Choice." Harvard University Mimeo.

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

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

[31] 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.

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

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

Notes

[1]Our experimental design is, of course, less suitable to answering questions about the dynamics of opinion change on particular substantive issues than are the research designs of more traditional social psychological experiments. As is usually the case, some degree of tradeoff between empirical verisimilitude and experimental control — i.e., between external validity and internal validity — is unavoidable. In the long run, both styles of research can offer distinctive yet complementary insights into the nature of deliberation.

[2]The strategic deliberative environment we analyze is most closely related to the models by Hafer and Landa (2005, 2007), who also consider the deliberation with arguments that are provable to some but not all members of the audience, but do not analyze collective decision-making that is our focus below. Other related theoretical models include Lipman and Seppi (1995) and Glazer and Rubinstein (2005), which analyze messages with similar informational content but assume that they are equally persuasive to all.

[3]The supplemental appendix for reviewers contains the instructions given to subjects, demonstrating the way in which the experimental scenario was presented in the laboratory.

[4]Later, after presenting some experimental results, we proceed to consider explicitly other, non-Bayesian cognitive-behavioral ideal types to provide an account of the microfoundations of deliberation that is more consistent with the observations in our experimental data.

[5]Formal arguments that these strategies are weakly dominant are contained in the reviewers' supplemental appendix that we will make available on-line in the event of publication.

[6]Note that we also instantiated configurations CCD and BCD, but because these are strategically equivalent to ABB and ABC respectively, we subsume CCD into references to ABB and BCD into references to ABC except where specifically noted. Our choice of deliberative situations allows us to observe and control for a variety of features that are strategically irrelevant in our model but which could plausibly affect subject behavior – such as whether or not subjects know their true numbers for certain (when this would not actually affect their optimal strategies) or whether subjects behave differently when their active fragments are on the left (lower) or on the right (higher) –

thereby minimizing potential confounds in interpretation. Because only agents with adjacent active fragments can potentially persuade one another, configurations such as ABD or ACD involve at least one player's deliberative behavior being inconsequential; we exclude such groupings because they are of more limited interest.

[7]Individual decisions are the unit of analysis throughout the paper unless otherwise noted.

[8]And four of the six instances in which a moderate spoke during these periods came from a single subject. Note that the corresponding figure over all 30 rounds is 94.9% (501/528).

[9]The question was: "Suppose that the set of possible true numbers is 26,68,89. Suppose that the frequency of the true number 26 is 35%, that the frequency of the true number 68 is 45%; and that the frequency of the true number 89 is 20%. And suppose that a person in your group is told that 6 is a fragment of his or her true number. What is his or her most likely true number given this information?" As feedback, subjects saw their own answer and the text: "The correct answer was 68. If the person has a fragment 6, his or her true number must be either 26 or 68 because 89 does not contain the fragment 6. And 68 occurs with greater frequency than 26."

[10]Note for now that, when asked "Did you find the problem at hand difficult or easy?" and "Do you think other people found the problem difficult or easy?" only 4 of the 36 subjects (11.1%) gave responses indicating a belief that they found it easier than their counterparts, and only one of these 4 subjects exhibits deliberative and voting behavior that is consistent with the alternative explanation. As such, these responses weigh strongly against an interpretation of the data in which Bayesian subjects always attempt to activate the latent fragments of other subjects whose capacities are believed to be much lower. See also fn. 16 below.

[11]Some recent economics literature usefully interprets experimental findings using a similar approach of defining a hierarchy of cognitive types - e.g. Camerer, Ho, and Chong (2004); Costa-Gomes, Crawford, and Broseta (2001).

[12]The nature of the following discussion, and even the specific classifications, would be practically unchanged using data from all periods. As noted, the Table 3 regression contains no evidence of aggregate learning in Periods 13-30. Additional regression specifications attempted to isolate likely

points at which experience-based learning might have occurred. For example, a subject choosing to listen as a moderate who receives no message might wonder why, and see that there could be a strategic advantage to listening as an extremist – thus discovering Bayesian reasoning. However, regression specifications noting subjects' exposure to such experiences, either period-to-period or with greater lag, all returned insignificant learning effects.

[13]Note that this is a conservative standard that cuts against the affirmative classifications of the Watsonian and the Unreflective Bayesian types.

[14]However, given the small sample size of situations in which 'B' agents received only a foreign fragment in the Listening Case (ABB), it is not possible to make a strong statistical claim that this correlation did not arise due to chance. Obtaining the necessary sample size to make robust statistical inferences is difficult for several reasons. In order to receive a foreign fragment in the Listening Case (ABB), it must be the case that a given moderate agent chooses to listen; that the extreme agent chooses to speak; and that the extreme agent's active argument is not part of the moderate agent's true number (that is, the moderate agent's true number is $\overline{BC}$). But in the Listening Case (ABB), the moderate agent's true number is more likely actually to be $\overline{AB}$ than it is to be $\overline{BC}$ – so instances in which the moderate agent receives a foreign fragment rather than the fragment A are comparatively infrequent. And, of course, at least from the Bayesian perspective, it is out-of-equilibrium for the extreme agent to choose to speak in such a setting in the first place.

[15]Our post-experimental survey asked subjects to explain how they chose to speak vs listen; whether communication was helpful in deciding how to vote and if so, how; and how difficult they thought the problem posed was for them, and for others. We blind-coded subjects' responses into ideal types according to the type of reasoning they suggested. These codings yielded a very high correlation with actual behavior in the experiment, providing further support for our interpretation of the results. (All 4 subjects classified as Bayesian [or Leaning] gave consistently Bayesian survey responses. Of the 20 subjects classified as Watsonian [or Leaning]/Unreflective Bayesian, 14 gave consistently Watsonian responses, 4 gave Watsonian responses to the deliberation questions but Bayesian answers to the voting question, and 2 gave responses of mixed character.)

Table 1a. Weakly Dominant Deliberative Strategies for Bayesian Agents

Deliberative Setting	A active (Extreme)	B active (Moderate)	C active (Extreme)
ABB, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	Speak	Listen	-
ABB, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	**Listen**	Listen	-
ABC, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	Speak	Listen	**Listen**
ABC, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	**Listen**	Listen	Speak

Note: The weakly dominant deliberative strategies for agents with an 'A' or 'C' active fragment are indicated in plain text for the Speaking Case situations, and in boldface for the Listening Case situations.

Table 1b. Dominant Voting Strategies for Bayesian Agents with a 'B' (Moderate) Active Fragment, by Communications Received

Deliberative Setting	nothing	foreign fragment only	latent fragment
ABB, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	$\overline{BC}$	$\overline{BC}$	actual true number
ABB, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	$\overline{AB}$	$\overline{\mathbf{BC}}$	actual true number
ABC, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	$\overline{BC}$	$\overline{BC}$	actual true number
ABC, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	$\overline{AB}$	$\overline{AB}$	actual true number

Note: subjects with an 'A' ('C') active fragment always have a dominant strategy to vote for $\overline{AB}$ *over* $\overline{BC}$ *(*$\overline{BC}$ *over* $\overline{AB}$*).*

Table 2a. Aggregate Communications Behavior (Periods 13-30)

Deliberative Setting	A active (Extreme)	B active (Moderate)	C active (Extreme)
ABB, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	79.2% Speak (38/48)	99.0% Listen (95/96)	-
ABB, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	**54.2% Speak (26/48)**	97.9% Listen (94/96)	-
ABC, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	78.3% Speak (47/60)	96.7% Listen (58/60)	**58.3% Speak (35/60)**
ABC, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	**66.7% Speak (40/60)**	98.3% Listen (59/60)	73.3% Speak (44/60)

Table 2b. Aggregate Communications Behavior When Subjects Have 'A' or 'C' (Extreme) Active Fragments

.	Speak (S Case)	Listen (S Case)	Speak (L Case)	Listen (L Case)
Experimental Data				
All Periods	196 (71.0%)	80 (29.0%)	169 (61.2%)	107 (38.8%)
Periods 1-12	67 (62.0%)	41 (38.0%)	68 (63.0%)	40 (37.0%)
Periods 13-30	129 (76.8%)	39 (23.2%)	101 (60.1%)	67 (39.9%)
Theoretically Expected Behavior				
Bayesian	100%	0%	0%	100%

L Case - Listening Case; S Case - Speaking Case.

Table 2c. Aggregate Voting Behavior of Subjects with a 'B' (Moderate) Active Fragment (All Periods)

Deliberative Setting	nothing	foreign fragment only	latent fragment
ABB, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	91.7% $\overline{BC}$ (44/48)	91.8% $\overline{BC}$ (67/73)	92.7% actual true number (38/41)
ABB, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	94.6% $\overline{AB}$ (70/74)	**56.3% $\overline{BC}$ (18/32)**	100% actual true number (59/59)
ABC, $Pr(\overline{BC}\|B) > Pr(\overline{AB}\|B)$	100% $\overline{BC}$ (18/18)	95.5% $\overline{BC}$ (21/22)	96.4%actual true number (54/56)
ABC, $Pr(\overline{BC}\|B) < Pr(\overline{AB}\|B)$	100% $\overline{AB}$ (18/18)	85.7% $\overline{AB}$ (12/14)	90.6%actual true number (58/64)

Note: subjects with an 'A' active fragment voted for $\overline{AB}$, their dominant strategy, 96.7% of the time (348/360). Subjects with a 'C' active fragment voted for $\overline{BC}$, their dominant strategy, 96.4% of the time (185/192).

Table 3. Factors Affecting Deliberative Choice in Subjects with 'A' or 'C' (Extreme) Active Fragments (Periods 13-30)

Probit Regression with Robust Standard Errors

Dependent Variable: listen = 1 if subject listens, = 0 if subject speaks

$N = 336$; $PseudoR^2 = 0.0321$

listen	coefficient	robust SE	z	$P > \|z\|$
lnperiod	-.1352728	.3447588	-0.39	0.695
caselisten	.4685002**	.1472395	3.18	0.001
ABB (not ABC)	.0563819	.1722589	0.33	0.743
right-handed	.2296582	.1752874	1.31	0.190
xdist	.0100337	.0216525	0.46	0.643
xfrac	-.4254516	1.569162	-0.27	0.786
pdist	-1.562123	2.172081	-0.72	0.472
pfrac	.7623341	1.506065	0.51	0.613
constant	-.4361045	1.111846	-0.39	0.695

lnperiod is the natural log of the period number. **caselisten** is 1 for the Listening Case, 0 for the Speaking Case. **ABB (not ABC)** is 1 for ABB or CCD, 0 for ABC or BCD. **right-handed** is 1 for BCD or CCD, 0 for ABB or ABC. **xdist** is $|y_1 - y_2|$, the distance between the alternatives to be voted on. **xfrac** is $\frac{|y_1-y_2|}{CD-AB}$, the distance between the alternatives to be voted on relative to the distance between the most extreme true numbers. **pdist** is the absolute value of the difference of the unconditional probabilities associated with the alternatives to be voted on. **pfrac** is **pdist** divided by the sum of the unconditional probabilities associated with the alternatives to be voted on.

Table 4. Individual Communications Behavior When Subjects Have 'A' or 'C' (Extreme) Active Fragments (Periods 13-30)

Subject No.	Speak (S Case)	Listen (S Case)	Speak (L Case)	Listen (L Case)	Classification
1	0	3	0	5	Dev
2	4	1	1	2	Unc#
3	6	0	3	3	W/UB#
4	4	1	2	1	W/UB
5	5	0	3	0	W/UB*
6	3	0	0	5	B*
7	1	2	2	3	Dev*
8	6	0	0	6	B
9	2	4	2	4	Dev
10	1	5	1	5	Dev#
11	4	2	1	5	Lean B**
12	5	0	3	0	W/UB#
13	2	4	6	0	Dev
14	3	0	5	0	W/UB#
15	3	0	3	2	W/UB
16	5	0	3	0	W/UB*
17	2	1	3	2	Dev
18	5	0	3	0	W/UB
19	3	0	5	0	W/UB*
20	5	0	3	0	W/UB*
21	5	1	5	1	W/UB*
22	2	3	0	3	Dev
23	5	0	3	0	W/UB###
24	3	0	5	0	W/UB**
25	3	0	5	0	W/UB*
26	6	0	3	3	W/UB*
27	6	0	0	6	B*
28	6	0	6	0	W/UB#
29	6	0	4	2	W/UB*
30	5	0	1	2	Unc*
31	4	2	6	0	Lean W/UB
32	1	2	4	1	Dev*
33	0	3	0	5	Dev###
34	2	3	2	1	Dev##
35	2	1	5	0	W/UB
36	4	1	3	0	W/UB#*
TOTALS	129 (76.8%)	39 (23.2%)	101 (60.1%)	67 (39.9%)	.

Dev - Deviant; Unc - Unclassified; W - Watsonian; UB- Unreflective Bayesian; B - Bayesian; Lean B - Leaning Bayesian; Lean W - Leaning Watsonian; L Case - Listening Case; S Case - Speaking Case.

The votes of those subjects who in any period of the experiment received only a foreign fragment while possessing active fragment B in the Speaking Case (ABB) are indicated next to their classification type. Each occurrence of (#) denotes one incorrect vote for the prior while each occurrence of (*) denotes one correct vote against the prior in such a situation.