

# Supplemental Appendix for “Cognition and Strategy: A Deliberation Experiment” by Dickson, Hafer, and Landa

In Appendix A below, we present the formal analysis containing proofs for the equilibrium behavioral predictions in the game analyzed in the main manuscript. The second part of this document, Appendix B, contains our instructions to subjects in the experiment. Finally, Appendix C addresses points referred to in two footnotes in the main manuscript.

## 1 Appendix A: Formal analysis

In this appendix, we consider the incentives perceived by Bayesian and by Overspeaking Watsonian agents in a variety of deliberative settings. The scope of the analysis covers all of the situations faced by our experimental subjects. The solution concept that we employ is weak dominance.

Let  $p(t)$  be the unconditional probability of type  $t$ . Member  $i$ 's prior beliefs about  $t_i$  are described by  $p(t|a_i)$ , which is derived from  $p(t)$  and  $\Pr(a_i|t) = \begin{cases} \frac{1}{2} & \text{if } a_i \in \{t^1, t^2\} \\ 0 & \text{otherwise} \end{cases}$  using Bayes Rule. Let  $M_i$  be the set of messages received by  $i$  during deliberation; then  $i$ 's posterior belief  $p(t|a_i, M_i, \lambda)$  is also derived from  $p(t|a_i)$  using Bayes Rule. Given that the mapping of type to ideal point is one-to-one and commonly known, in the interests of simplicity of reference, and with some abuse of notation, we use type and ideal point interchangeably (e.g.,  $p(\overline{AB})$  in place of  $p((A, B))$ ).

Let  $\mathcal{M}_i(a)$  be the set of all possible  $M_i$  given  $a$ . Note that  $M_i$  could contain two identical elements if  $i$  receives two identical messages. (The identity of the sender and the receiver is not observable.) We abuse set notation by writing, e.g.,  $M_i = \{B, B\}$  is  $i$  receives two “ $B$ ” messages.

Let  $v^*(p(\cdot))$  be  $i$ 's weakly dominant voting strategy. Given that  $p(\overline{AB}|A) = 1$  and  $p(\overline{CD}|B) = p(\overline{AB}|C) = 0$ ,

$$v_i^*(p(\cdot)) = \begin{cases} \overline{AB} & \text{if } p(\overline{AB}|a_i, M_i, \lambda) > p(\overline{BC}|a_i, M_i, \lambda) \\ \overline{BC} & \text{if } p(\overline{BC}|a_i, M_i, \lambda) > p(\overline{AB}|a_i, M_i, \lambda). \end{cases} \quad (1)$$

Let  $\lambda_i^{BR} : \{0, 1\}^2 \rightarrow 2^{\{0,1\}}$  be  $i$ 's best response to  $\lambda_{-i}$  given  $v^*$ , i.e.,

$$\lambda_i^{BR} = \arg \max E[u_i(\lambda_i, \lambda_{-i}, v^*, a, p(t))].$$

## 1.1 Equilibrium Behavior of Bayesian Agents

**Situation  $ABB$ , Vote  $\overline{AB}$  vs  $\overline{BC}$ .**  $\mathcal{M}_1((A, B, B)) = \{\emptyset, \{B\}, \{B, B\}\}$ ;

$\mathcal{M}_2((A, B, B)) = \mathcal{M}_3((A, B, B)) = \{\emptyset, \{A\}, \{\text{foreign}\}, \{B\}, \{A, B\}, \{B, \text{foreign}\}\}$ .

Suppose without loss of generality that  $a_1 = A$ ,  $a_2 = a_3 = B$ . From (1), if  $p(\overline{BC}|B) > p(\overline{AB}|B)$ , then  $\forall j \in \{2, 3\}$ , we have (1)  $\forall M_j$  s.t.  $A \notin M_j$ ,  $v_j^*(p(\cdot|B, M_j, \lambda)) = \overline{BC}$ ; and (2)  $\forall M_j \ni A$ ,  $v_j^*(p(\cdot|B, M_j, \lambda)) = \overline{AB}$ . If  $p(\overline{BC}|A) < p(\overline{AB}|B)$ , then  $\forall j \in \{2, 3\}$ , we have (1)  $\forall M_j \ni \text{"foreign"}$ ,  $v_j^*(p(\cdot|B, M_j, \lambda)) = \overline{BC}$ ; and (2)  $\forall M_j$  s.t. "foreign"  $\notin M_j$ ,  $v_j^*(p(\cdot|B, M_j, \lambda)) = \overline{AB}$ . Thus,

$$\lambda_1^{BR} = \begin{cases} \{0\} & \text{if } p(\overline{BC}) > p(\overline{AB}) \text{ and } \exists j \in \{2, 3\} \text{ s.t. } \lambda_j = 1 \\ \{1\} & \text{if } p(\overline{BC}) < p(\overline{AB}) \text{ and } \exists j \in \{2, 3\} \text{ s.t. } \lambda_j = 1 \\ \{0, 1\} & \text{if } \lambda_2 = \lambda_3 = 0 \end{cases}$$

$$\lambda_2^{BR} = \lambda_3^{BR} = \begin{cases} \{1\} & \text{if } \lambda_1 = 0 \\ \{0, 1\} & \text{if } \lambda_1 = 1, \end{cases}$$

and the unique weakly dominant strategies are

$$\lambda_1^* = \begin{cases} \{0\} & \text{if } p(\overline{BC}) > p(\overline{AB}) \\ \{1\} & \text{if } p(\overline{BC}) < p(\overline{AB}) \end{cases}$$

$$\lambda_2^* = \lambda_3^* = 1. \blacksquare$$

**Situation  $ABC$ , Vote  $\overline{AB}$  vs  $\overline{BC}$ .**  $\mathcal{M}_1((A, B, C)) = \{\emptyset, \{B\}, \{\text{foreign}\}, \{B, \text{foreign}\}\}$ ;  $\mathcal{M}_2((A, B, C)) = \{\emptyset, \{A\}, \{\text{foreign}\}, \{C\}, \{A, \text{foreign}\}, \{C, \text{foreign}\}\}$ ;  $\mathcal{M}_3((A, B, C)) = \{\emptyset, \{B\}, \{\text{foreign}\}, \{B, \text{foreign}\},$

{foreign,foreign}}.

Suppose without loss of generality that  $a_i = A, a_2 = B, a_3 = C$ . From (1),  $v_1^*(p(\cdot)) = \overline{AB}$  and  $v_3^*(p(\cdot)) = \overline{BC} \forall M, \forall \lambda$ . Thus,  $\lambda_2^* = 1$  is a unique weakly dominant strategy.

If  $p(\overline{BC}|B) > p(\overline{AB}|B)$ , then  $v_2^*(p(\cdot|B, M_2, \lambda)) = \overline{AB}$  iff  $M_2 \ni A$  or both  $M_2 = \text{“foreign”}$  and  $\lambda_1 = 1, \lambda_3 = 0$ ; else  $v_2^*(p(\cdot|B, M_2, \lambda)) = \overline{BC}$ . Thus  $\lambda_3^* = 1$  is a unique weakly dominant strategy.  $\lambda_1^{BR}(1, 1) = \{0\}$ , thus  $\lambda_1^* = 0$ , given  $\lambda_2^* = \lambda_3^* = 1$ .

If  $p(\overline{AB}|B) > p(\overline{BC}|B)$ , then  $v_2^*(p(\cdot|B, M_2, \lambda)) = \overline{BC}$  iff  $M_2 \ni C$  or both  $M_2 = \text{“foreign”}$  and  $\lambda_1 = 0, \lambda_3 = 1$ ; else  $v_2^*(p(\cdot|B, M_2, \lambda)) = \overline{AB}$ . Thus  $\lambda_1^* = 1$  is a unique weakly dominant strategy.  $\lambda_3^{BR}(1, 1) = \{0\}$ , thus  $\lambda_3^* = 0$ , given  $\lambda_1^* = \lambda_2^* = 1$ . ■

## 1.2 Equilibrium Behavior of Overspeaking Watsonian Agents

Note that the arguments identical to the ones below establish the identical optimal speaking/listening choices of the Overspeaking Bayesian agents.

**Situation  $ABB$ , Vote  $\overline{AB}$  vs  $\overline{BC}$ .** Suppose without loss of generality that  $a_i = A, a_2 = a_3 = B$ . From (1),  $v^*(p(\cdot|B, M_1, \lambda)) = \overline{AB} \forall M_1, \forall \lambda$ . Because  $v^*(p(\cdot|B, M_j, \lambda)) = v^*(p(\cdot|B, M_j \setminus B, \lambda)) \forall M_j, \forall \lambda$ ,  $\lambda_2^* = \lambda_3^* = 1$  is the unique weakly dominant strategy.

If  $p(\overline{BC}|A) < p(\overline{AB}|B)$ , then  $\forall j \in \{2, 3\}$ ,  $v^*(p(\cdot|B, M_j, \lambda)) = \overline{AB} \forall M_j, \forall \lambda$ . Thus  $\lambda_1^{BR}(1, 1) = \{0, 1\}$ , and  $\lambda_1^* \in \{0, 1\}$ .

If  $p(\overline{BC}|A) > p(\overline{AB}|B)$ , then  $\forall j \in \{2, 3\}$ ,  $v^*(p(\cdot|B, M_j, \lambda)) = \overline{AB}$  iff  $M_j \ni A$ ; else  $v^*(p(\cdot|B, M_j, \lambda)) = \overline{BC}$ . Thus,  $\lambda_1^* = 0$  is the unique weakly dominant strategy. ■

**Situation  $ABC$ , Vote  $\overline{AB}$  vs  $\overline{BC}$ .** Suppose without loss of generality that  $a_i = A, a_2 = B, a_3 = C$ . From (1),  $v^*(p(\cdot|A, M_1, \lambda)) = \overline{AB} \forall M_1, \forall \lambda$  and  $v^*(p(\cdot|C, M_3, \lambda)) = \overline{BC} \forall M_3, \forall \lambda$ . Thus,  $\lambda_2^* = 1$  is a unique weakly dominant strategy.

If  $p(\overline{AB}|B) > p(\overline{BC}|B)$ , then  $v^*(p(\cdot|B, M_2, \lambda)) = \overline{BC}$  iff  $M_2 \ni C$  else  $v^*(p(\cdot|B, M_2, \lambda)) = \overline{AB}$ . Thus  $\lambda_3^* = 0$  is a unique weakly dominant strategy.  $\lambda_1^{BR}(1, 0) = \{0, 1\}$ , so  $\lambda_1^* \in \{0, 1\}$ .

If  $p(\overline{BC}|B) > p(\overline{AB}|B)$ , then  $v^*(p(\cdot|B, M_2, \lambda)) = \overline{AB}$  iff  $M_2 \ni A$  else  $v^*(p(\cdot|B, M_2, \lambda)) = \overline{BC}$ . Thus  $\lambda_1^* = 0$  is a unique weakly dominant strategy.  $\lambda_3^{BR}(0, 1) = \{0, 1\}$ , thus  $\lambda_3^* \in \{0, 1\}$ . ■

## 2 Appendix B: Instructions to Subjects

### Introduction

This is an experiment on decision-making. In the following experiment you will make a series of choices. At the end of the experiment, you will be paid depending on the specific choices that you made during the experiment and the specific choices made by other people. If you follow the instructions and make appropriate decisions, you may make an appreciable amount of money.

The experiment will consist of 30 different rounds. At the beginning of each round of the experiment, you will be randomly assigned into a group of three people (that is, you will be randomly grouped with two other people in the lab). As such, the composition of groups will be different in each round of the experiment. All of your interactions with others will take place anonymously through a computer terminal, so that your personal identity will never be revealed to others and you will not know who is in your group. In each round, each member of your group will have what we call a true number, and will be given partial information about it. Each member of your group will, then, have an opportunity to communicate with the other members. At the end of each round, each group member will be asked to cast a vote in an election between two different two-digit numbers. Ultimately, you will make the most money if the number that wins the majority of votes is as close as possible to your true number.

The structure of the experiment is now described in greater detail.

#### (1) Initial Information

At the beginning of each round of the experiment, you will receive four kinds of information. First, you will be given the set of possible true numbers. Such a set might look as follows: {24, 47, 79}. One of these numbers will be your own true number, although you will not be told which of these numbers it is. Similarly, each of the other members of your group will also have a true number that comes from this set, and will be shown the set of numbers but will not be told what his or her particular true number is. It may be the case that every member of your group has a different true number, but it may also be that some of the members of your group have the same true number.

Thus, the true number of any other member of your group may be the same as your true number, or it may be different. A second piece of information that is given to everyone in your group will further help you figure out your true number. For each of the possible true numbers, you will be told how likely it would be, if you didn't know anything else about what your true number was, that a given number in the set of true numbers would, in fact, turn out to be your true number. For example, you may be told that that likelihood is 20% for number 24, 50 % for number 47, and 30% for number 79.

But, in fact, we will give you an additional hint. To help you further improve your guesses of your own true numbers, each of you will receive a third piece of information a one-digit fragment of your own true number. For example, if your true number were 47, you might be given the fragment 4 or the fragment 7. At the same time that you receive this fragment of your own true number, you will also be told the set of one-digit fragments received by all the other members of your group. Finally, you will be told which alternatives will be voted on in the election held at the end of the round. For example, you may be told that the election will be between the number 24 and the number 47.

## (2) Communication

After you have received the initial information, all of you will have the opportunity to communicate with other members of your group. Communication in your group will take a particular form. All members of your group will, simultaneously, be asked to choose either to (a) send their respective fragments to all the other group members at once; or to (b) receive the fragments that are being sent by all those members of your group who choose to send their fragments (that is, by all the members of your group who choose the option (a)). Each member of the group must choose either to send or to receive; it is not possible to do both.

If you choose to receive fragments, you will receive fragments from those and only those members of your group who choose to send their fragments. Regardless of how many fragments you may receive, you will only be able to see a fragment sent to you if it is part of your true number i.e., if it is a fragment that completes your true number or a fragment you already have. If you receive other

fragments that is, fragments that are not part of your true number you will only be told that you have received foreign fragments, and how many of them. For example, suppose that your fragment is 4 and your true number is 47. If you choose to receive fragments, and a member of your group sends the fragment 7, you will be told that you have received the fragment 7. Similarly, if you choose to receive fragments, and a member of your group sends the fragment 4, you will be told that you have received the fragment 4. However, if you choose to receive fragments, and a member of your group sends the fragment 9, you will simply be told that you have received a foreign fragment, because 9 is not a fragment of 47. Likewise, every other member of your group who chooses to receive fragments will see only those fragments that are a part of his or her own true number, and will be told how many foreign fragments they have received in the event that any such fragments are received.

If you choose to send your fragment, it will be received by those and only those members of your group who choose to receive fragments, but you yourself will receive no fragments of any kind. Similarly, if other members of your group choose to send their respective fragments, these fragments will be received by those and only those members who choose to receive, but the sending members themselves will receive no fragments of any kind.

Communication may then have two effects: First, you might learn something about your own true number that could help you determine how you will vote. And second, you might cause other members of your group to change their guesses about their own true numbers and so affect their votes. Always remember that you will make more money the closer the number that wins the majority of votes is to your true number.

### (3) Election and Payoffs

Once the communication phase is complete, all members of your group will be asked to vote in the election that was specified earlier in the example above, this was 24 vs. 47. The number that receives the majority of votes wins. Because there are three members in each group (including you), there cannot be a tie. The winning number from the election will be used to determine the payoffs for the round.

Your payoffs for each round will be calculated as follows. If the winning number matches your true number exactly, you will receive 80 cents. If the winning number does not match your true number exactly, you will receive 80 cents less 1 cent for each unit of distance between the winning number and your true number. For example, if your true number is 62, but the winning number is 12, there are 50 units of distance between your true number and the winning number, and you would receive  $(80-50)$  cents, that is, 30 cents.

The experiment will consist of 30 rounds like the one just described. In each of these rounds, you will be assigned a new true number; receive initial information about the new set of possible true numbers; be told what numbers are to be voted upon; be given a fragment of your true number; have an opportunity to attempt communication with other members of your group; and, along with your fellow group members, cast a vote. Your total payoff for the experiment is the sum of your payoffs from each of these rounds, plus the show-up fee. Remember, your payoff in each round is higher, the closer the winning number is to your true number.

### 3 Appendix C: Discussion relevant to Footnotes 13 and 21

Here we provide evidence rejecting an alternative interpretation of our findings that might at first blush seem consistent with the evidence of overspeaking in Section 5.1. Suppose that a population of agents exists that has no or a limited understanding of the unconditional probabilities associated with the true numbers. As voters, such agents may have little idea how to cast their ballots when their initial active fragment is moderate and when their latent fragment is not activated during deliberation; perhaps they would simply vote randomly. However, these agents would be highly likely to vote correctly if their latent fragment *were* activated – with both fragments known, the probability information would become irrelevant. If such agents were present in sufficient numbers, a Bayesian with an extreme active fragment might have an incentive to speak – even in the Listening Case – if she believed that the likelihood of persuading such agents exceeded the risk of alienating Bayesian listeners.

However, this alternative account is implausible for several reasons. First, subjects overwhelmingly gave correct answers to our pre-experiment quiz question about the unconditional probabilities (34/36, 94.4%).<sup>1</sup> Second, as noted in Conclusion 4, subjects with a moderate active fragment who receive no signal at all practically always vote in a way consistent with the unconditional probability information, further suggesting that very few if any subjects possess the particular deficiency in understanding described above that might give Bayesian agents incentives always to speak.

Finally, subjects’ responses to our post-experiment survey provide further evidence that few subjects imagined others to understand the scenario more poorly than they themselves did. 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.

Our post-experimental survey also 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] in the data gave survey responses consistent with Bayesian reasoning. Of the 20 subjects classified as Overspeaking Watsonian [or Leaning]/Overspeaking Bayesian, 14 gave survey responses consistent with Overspeaking Watsonian reasoning, 4 gave responses consistent with Overspeaking Bayesian reasoning, and 2 gave responses that defied straightforward classification.)