Effective Persuasion[*]

Ying Chen[†] Wojciech Olszewski[‡]

December 30, 2012

Abstract

Do elementary statistics or equilibrium theory deliver any rules of thumb regarding how we should argue in debates? We provide an answer in a model in which each discussant wants the audience to believe that the actual state is his favorite state. We show that if the discussants' payoffs in the audience's posterior are concave above the prior, convex below the prior and exhibit loss aversion, then the leading discussant should give precedence to the weaker argument, and the following discussant should respond to a weak argument with a weak argument, and to a strong argument with a strong argument. Under similar preferences, when choosing between independent and correlated arguments, the leading discussant should give precedence to an argument potentially correlated with the follower's argument, and the following discussant should respond to a correlated argument with an independent argument, and to an independent argument with a correlated argument.

Keywords: *persuasion, debates, hard evidence, weak and strong arguments, correlated and independent arguments*

[*] We thank Joel Sobel for insightful comments and Mikhail Safronov for excellent research assistance. We are also grateful to Hanming Fang (the editor) and two anonymous referees for helpful suggestions and comments. Wojciech Olszewski thanks the National Science Foundation for research support (CAREER award SES-0644930).

[†] Department of Economics, Arizona State University, P.O. Box 879801, Tempe, AZ 85287-9801

[‡] Department of Economics, Northwestern University, 2001 Sheridan Rd. Evanston IL 60208-2600

1 Introduction

The way we argue is important for achieving our goals in a debate. After every debate in presidential or parliamentary elections, we often ask who "won" the debate. Other examples abound, ranging from deliberation contests in colleges to our everyday experience. Everyone who engages in a debate faces the problem of how to persuade effectively: Which points to raise and in which order? Which of them to emphasize and which of them to disregard? How to respond to the points raised by opponents?

Traditional models of communication, such as Crawford and Sobel (1982), focus on information transmission and do not discuss how the agents argue and what constitute good deliberation skills. These issues seem to be complex. Deliberation depends on what the discussants know about the audience and the opponents, as well as on the discussants' experience acquired in the process of learning by doing. Discussants may also attempt to sway the audience by stirring emotions and by exploiting other psychological effects.

In this paper, we propose a normative framework which assumes away the psychological side of debates. Our goal is to see whether formal models, based on Bayes' rule and equilibrium concepts, deliver any rules of thumb regarding the way we should argue in debates.

The optimal way to persuade is driven by the audience's expectations to a great extent. Indeed, suppose that we plan to make a certain point or present a certain piece of evidence, but the audience expects us to comment on another aspect, or present another piece of evidence. Then, when we make the point as intended, the audience may begin to think that we have little to say on the other aspect, or we lack the evidence they anticipated us to reveal, thus undermining the force of our argument. So, how effective a point really is depends on what we lose by not following the audience's expectation.

It is largely an empirical question what an audience's expectations are in a particular situation. Based on our theoretical study, we can say little about them. But we can say more about what the discussants want these expectations to be. We take the perspective that the discussants can shape the audience's expectations, or build reputation for arguing (presenting their evidence) in a certain manner. So, the question addressed in this paper is what reputation the discussants may, or should want to build. In this sense, the analysis is more applicable to repeated than to one-shot interactions.

More precisely, we study a model with two discussants and an audience. The discussants have conflicting objectives: each of them wants to convince the audience that the state of the world is the one he likes most, independent of the actual state. Each discussant has a finite number of pieces of hard evidence. Discussants move sequentially, and each of them can reveal at most one piece of evidence at a time. A positive probability of termination makes it possible that discussants will not manage to reveal all the evidence they have. Therefore, the order in which they reveal their evidence is important. Each discussant can commit to a strategy, that is, which piece of evidence he will reveal, contingent on the evidence he has and the evidence that has been revealed earlier.

For simplicity, we restrict attention to situations in which each discussant has at most two pieces of hard evidence and the debate terminates after each discussant presents at most one piece. This also allow us to minimize the controversies regarding the predictive power of equilibrium analysis since the games we study are simple. The optimal strategy of the discussant who moves second, called *the follower*, is derived from Bayes' rule, and deriving the optimal strategy of the discussant who moves first, called *the leader*, requires predicting the response of the follower in addition.

Within this simple model, we address two questions. Should discussants always present their strongest evidence first? And should they lean towards presenting the evidence that is independent or the evidence that is correlated with some evidence that their opponents potentially have?

The answers depend on the discussants' payoffs, which are functions of the audience's posterior belief. Because of the martingale property of information revelation, the discussants' strategies do not affect the audience's expected posterior, but affect only the *dispersion* of the posterior. Consequently, which strategy is optimal for a discussant depends on how much dispersion it creates in the audience's posterior and the discussant's risk attitude in the relevant region in which the posteriors lie.

For example, suppose the follower's payoff function is concave above the audience's prior, convex below the prior, and exhibits loss aversion (Kahneman and Tversky, 1979). Since a stronger piece of evidence is always more informative than a weaker piece of evidence, committing to revealing the stronger evidence first creates more dispersion in the audience's posterior. Combined with the discussant's risk attitude in the relevant regions, this implies that the follower should respond with weaker evidence to weaker evidence presented by the leader, and with stronger evidence to stronger evidence. Similarly, if the leader's payoff function is also concave above the prior, convex below the prior, and exhibits loss aversion (call this the Kahneman and Tversky preference), then he should give precedence to the weaker evidence since it creates less dispersion in the audience's posterior.

As to whether the discussants should first present independent or correlated evidence, the same underlying mechanics are at work, but because of the conditional correlation between certain pieces of evidence, which evidence is more informative (and thus creates more dispersion in the audience's posterior) is now determined endogenously. For example, if the leader's strategy is to give precedence to the correlated evidence but presents the independent evidence instead, then this reveals that the leader does not have the correlated evidence and thus the follower is likely to possess some correlated evidence in favor of his claim. Especially when the correlation is high, showing that he has the correlated evidence does not provide much additional information and the independent evidence is more informative. Under the Kahneman and Tversky preference, the follower is risk averse in the relevant region (above the prior in this case), implying that he should respond to the independent evidence with correlated evidence. Using a similar logic, we derive the follower's best responses to other strategies and the evidence presented. Anticipating the follower's response, the leader prefers giving precedence to the correlated evidence, at least when the evidence is highly correlated with some evidence that the follower may have. In sections 4.1 and 4.2, we use these theoretical results to discuss the experimental findings in Glazer and Rubinstein (2001).

Our study is primarily motivated by everyday debates, for example, discussions among colleagues at departmental meetings. However, to illustrate how we interpret our model and what we aim to capture, we now discuss some advice on how to argue from easily accessible sources and some examples of public debates.

Many portals offer advice on how to argue.[1] One example is the following suggestion on a yahoo voices website: "Acknowledge good points by your opponent - You should give credit to your opponent when he presents an interesting perspective or a point that you cannot challenge. This adds a certain level of respect and courtesy to the discussion, and makes it all the more meaningful when you adamantly disagree with something else." We view this as advice for building reputation in a debate.

Another example is the following advice offered on appellate.net for lawyers arguing before the U.S. Supreme Court: "If your opponent's argument did not impress the judges, simply stand up and confidently tell the judges that unless the Court has questions, we will waive rebuttal." We interpret this as suggesting responding to a weak argument with a weak argument.

Presidential debates provide many examples of persuasion strategies. For example, in the first U.S. presidential debate of 2012, when President Obama raised the argument that independent studies showed the only way to meet Governor Romney's pledge of not adding to the deficit is by burdening middle-class families, Romney responded by saying, "There are six other studies that looked at the study you describe and say it's completely wrong." This can be viewed as an example of

[1] For example, wikihow, appellate.net, yahoo voices, and websites of schools of communication.

responding with a correlated argument. In the same debate, when Mr. Romney raised an objection that trickle-down government is not the right answer for America, President Obama did not respond to it directly despite the moderator Jim Lehrer asking him to do so. Instead, he elaborated on what needs to be done, beginning with an improvement of the education system. This can be broadly interpreted as an example of responding with an independent argument.

There are other examples of persuasion strategies in less structured settings. A well-known example is President Obama's refusal to respond to the demand from sections of the political right that he releases his long-form birth certificate. Since Obama had released earlier a legal form of proof of his birthplace, this may be interpreted as an attempt of building reputation for not responding to weak arguments. A similar example of building such a reputation is the strategy of the Texas gubernatorial candidate Ann Richards in 1990. She refused to answer the charge that she smoked marijuana in the 60's. Richards not only won that election, but the issue did not arise during her reelection campaign in 1994.

Related literature

Earlier work on persuasion games focuses on characterizing when self-interested parties reveal all of the verifiable information they have and when they fail to do so (for example, Milgrom, 1981, Milgrom and Roberts, 1986, and Shin, 1994). We instead consider situations in which the discussants are constrained to reveal a limited amount of evidence, and try to characterize how the discussants should argue.

One inspiration for our paper is the recent work of Glazer and Rubinstein (2001, 2004, 2006). These authors are also interested in optimal rules of persuasion, but they view a debate as a mechanism by which an uninformed decision maker extracts information from informed discussants. We restrict attention to a particular game, one that in our opinion resembles many debates, but allow players to commit to debate strategies (or to build reputation for debating in a particular manner).

Other related papers include those by Dziuda (2011), Kamenica and Gentzkow (2011), Olszewski (2004), Sher (2009, 2011) and Thordal-Le Quement (2010). These papers also study the choice of arguments or questions in the context of persuasion or information elicitation. One aspect in Kamenica and Gentzkow (2011) that is similar to our paper is the importance of the curvature of the sender's utility in the receiver's belief, but the two paper study different models and address different questions. We are interested in finding the best way to argue for two adversarial sides that engage in a sequential debate given an information structure whereas Kamenica and Gentzkow (2011) study the optimal information structure for a sender who tries to persuade a receiver. Thordal-Le

Quement (2010) has a result that says that an expert sometimes omits some favorable evidence even when he can present unlimited amount of evidence. Although this is somewhat related to our finding that a discussant may want to present weak arguments first, it arises for a different reason: in Thordal-Le Quement (2010), the expert suppresses favorable evidence to signal that he holds little but very consistent evidence.

2 Basic Model

There are two (a priori) equally likely states of nature, $\omega = a$ or b; two agents (discussants), A and B, and an audience. The agents argue in front of the audience that the state is a or b, respectively.

Each agent is equipped with at most two signals, or pieces of "hard" evidence, in favor of his claim: s_I and t_I, where $I = A$ or B. Agents move sequentially, presenting one argument at a time. Agent A (the leader) moves first, presenting one piece of evidence available to him (if he has any) according to his choice; agent B (the follower) moves second, also presenting one piece of evidence (if he has any) according to his choice. Then the discussion ends.[2] The agents are not allowed to be silent when they have an argument. We make this simplifying assumption because our interests are restricted to two specific issues (the choice between weak and strong arguments and the choice between correlated and independent arguments). Of course, to study other issues, one may consider a richer model in which agents are allowed to be silent. In addition, in our analysis of whether a discussant should present the strongest evidence first, one can interpret presenting the weakest possible (uninformative) evidence as the option of staying silent.

The audience forms a posterior belief μ about the state of nature. This belief is contingent on the presented arguments, and the strategies of the two agents which the audience correctly anticipates. The agents' preferences are monotone in the audience's posterior. That is, the utility of agent A, denoted by u_A, is an increasing function of the probability assigned by belief μ to state a, denoted by μ_a, and the utility of agent B, denoted by u_B, is an increasing function of the probability assigned by belief μ to state b, denoted by μ_b. This approach is inspired by Geanakoplos, Pearce and Stacchetti (1989). Agents are expected-utility maximizers.

One may argue that the agents' utility should depend only indirectly on the audience's beliefs, through the audience's actions. This is consistent with our model. To illustrate, consider the

[2] The results would not be affected if we assumed that the discussion ends only with a positive probability, and with the complementary probability the agents who has two pieces of evidence would have a chance to present the second piece.

following example of a utility function u_I $(I = A, B)$ derived from actions: $u_I(\mu_i) = (\mu_i)^3 + 3(\mu_i)^2(1-\mu_i)$. This arises in a situation when the audience consists of three members who have private information on their thresholds for choosing one alternative over the other and the outcome is determined by majority voting. Specifically, suppose a member votes for alternative a if and only if he believes the probability that the state is a exceeds the threshold t and the agents' prior over each member's threshold t is uniformly distributed on $[0,1]$ and independent of the others' thresholds. Then, for agent A, when the posterior that the state is a is μ_a, the probability that alternative a is chosen by the three-member audience is $(\mu_a)^3 + 3(\mu_a)^2(1-\mu_a)$. If agent A's utility is linear in the probability that a is chosen, then we can represent it by $u_A(\mu_a) = (\mu_a)^3 + 3(\mu_a)^2(1-\mu_a)$. Note that it is convex on $[0,1/2]$ and concave on $[1/2,1]$.[3] (Because of symmetry, B has a similar utility function: $u_B(\mu_b) = (\mu_b)^3 + 3(\mu_b)^2(1-\mu_b)$.)

2.1 Information Structure

The following table exhibits the prior distribution over signals, contingent on $\omega = a$.

	$\neg s_B$	s_B
s_A	$(1-\varepsilon)^2 + \rho\varepsilon(1-\varepsilon)$	$(1-\rho)\varepsilon(1-\varepsilon)$
$\neg s_A$	$(1-\rho)\varepsilon(1-\varepsilon)$	$\varepsilon^2 + \rho\varepsilon(1-\varepsilon)$

where $0 < \varepsilon < 1/2$, and $0 \leq \rho < 1$. The prior is symmetric contingent on $\omega = b$. That is, if the state is a, then it is more likely that agent A has signal s_A (which in the table is denoted simply by s_A) but agent B does not have signal s_B (this is denoted by $\neg s_B$) than that agent A does not have signal s_A but agent B has signal s_B. And if $\rho \neq 0$, signals s_A and $\neg s_B$ are (positively) correlated conditionally on the state of nature. Contingent on the state being a, the odds that agent B has signal s_B are ε, but contingent in addition on agent A having signal s_A, they are $(1-\rho)\varepsilon$.

Agent $I = A$ or B obtains signal t_I with probability $1-\delta$, where $\delta \leq 1/2$, contingent on $\omega = i$, and he obtains t_I only with probability δ, contingent on $\omega \neq i$. Signals t_A and t_B are conditionally independent, and they are conditionally independent of signals s_A and s_B.

The model easily generalizes to any finite number of agents, any finite number of signals, and more general prior probability distributions. Although this extension is promising for future research, the simpler version of the model is sufficient to derive our main results.

[3] More generally, if the audience consists of an odd number of n members and they vote by majority rule, then the utility function is $u_A(\mu_a) = \sum_{i=0,\ldots,(n+1)/2} C_n^i (\mu_a)^{n-i} (1-\mu_a)^i$. It is straightforward, although somewhat tedious, to show that this function is convex on $[0,1/2]$ and concave on $[1/2,1]$ (details are omitted).

2.2 Strategies and Equilibria

Let $r_I \in R_I$ denote a (pure) strategy of agent I $(I = A, B)$, a mapping from agent I's information set to an argument $e_I \in \{s_I, t_I, \emptyset\}$. Agent A's information set consists of the signals he has, and agent B's information set consists of the signal he has and the argument e_A presented by agent A.

Since each agent I must present an argument when he has (at least) one, he has only one decision to make: whether to present s_I or t_I, when he has both signals at hand. Agent B's decision may depend on the signal that has been revealed by agent A (or the lack thereof).

We analyze a game in which agents can commit to their strategies ex ante; that is, each agent makes a binding commitment to play in a certain way without knowing what signals he has. Agents commit to their strategies sequentially; agent A commits to his strategy first, and given the choice of agent A's strategy, agent B commits to his strategy. (We show in section 5.4 that our main results hold even if the agents commit to their strategies simultaneously.)

More precisely, consider a strategy profile (r_A, r_B). Together with the distribution of signals, (r_A, r_B) generates a distribution of audience posterior $F_i^{r_A,r_B}(\mu_i)$ where μ_i satisfies Bayes' rule whenever applicable.

Fix the leader's strategy r_A and let $r_B^*(r_A) \in \arg\max_{r_{B\in R_B}} \int u_B(\mu_b)\, dF_b^{r_A,r_B}$. The strategy $r_B^*(r_A)$ maximizes the follower's ex ante expected payoff when the leader plays r_A. Hence, $r_B^*(r_A)$ is a strategy that the follower would like to commit to, given that the leader plays r_A.

Similarly, let $r_A^* \in \arg\max_{r_{A\in R_A}} \int u_A(\mu_a)\, dF_a^{r_A,r_B^*(r_A)}$. Then r_A^* maximizes the leader's ex ante expected payoff, given that the follower responds by playing $r_B^*(r_A)$. Hence, r_A^* is a strategy that the leader would like to commit to.

As we show in our later analysis, the solution of the commitment game (r_A^*, r_B^*) is often also an equilibrium in the game without commitment.[4] An important advantage of analyzing the commitment game is that it allows us to provide unique predictions. In contrast, the game without commitment has a serious problem of multiple equilibria and does not generate sharp predictions, as we show in section 5.1. This commitment assumption can be motivated either by reputation effects, or by costly information acquisition. We postpone an extensive discussion of this assumption to section 5.2.

[4] The main difference is that an equilibrium in the commitment game requires the agents' strategies to be optimal ex ante, before they learn their signals, whereas in the game without commitment, an equilibrium requires the agents' strategies to be optimal after they learn their signals.

3 Weak Versus Strong Evidence

3.1 The Follower's Problem: Always Respond with the Strongest Argument Available?

Our first application addresses the following question: Should the follower respond with a weak argument to a weak argument, and with a strong argument to a strong argument? Or, should he always respond with his strongest argument? For example, suppose the opponent gives an unconvincing argument. Should one counter-argue (or build a reputation for counter-arguing) decisively, or rather disregard the opponent's argument, trying to make the impression that he could have given a powerful response, but does not want to get involved in a discussion of low quality?

Formally, suppose that $\rho = 0$, and $\varepsilon < \delta = 1/2$. So there is no conditional correlation between different signals, and signals s_A and s_B are stronger, that is, more informative about the state of the world, than t_A and t_B. For simplicity, we assume that t_A and t_B convey no information about the state of the world. (In Appendix B, we extend the model to incorporate the case that the weaker signal is also informative, that is, $\varepsilon < \delta < 1/2$.)

To illustrate how the audience's posteriors depend on the strategies and evidence presented, suppose first that the leader's strategy is to present the weak signal t_A when he has both t_A and s_A (we call this the "conciliatory" strategy). Suppose also that the leader has presented the weak evidence t_A in favor of his claim. The following table exhibits μ_b, the posterior belief of the audience that $\omega = b$ under each strategy of the follower, given the signals at the follower's disposal:

	only t_B	only s_B	both s_B & t_B
str. t_B	$\mu_b = 1/2$	$\mu_b = 1 - \varepsilon$	$\mu_b = 1/2$
str. s_B	$\mu_b = \varepsilon$	$\mu_b = 1 - \varepsilon$	$\mu_b = 1 - \varepsilon$
ex ante prob.	1/8	1/8	1/8

Table 1. The leader who plays the conciliatory strategy presented t_A.

The columns correspond to the following events: the follower has only signal t_B, only signal s_B, and both s_B and t_B, respectively. The event that the follower has no signal is omitted because the follower's strategy is irrelevant for the audience's posterior in this event. The first two rows correspond to the two strategies of the follower: str. t_B (str. s_B) is an abbreviation for the strategy of responding with t_B (s_B) if he has both signals. These are of course not complete descriptions of strategies as they have been defined in section 2.2, since str. t_B and str. s_B specify only what the agent does when he has both signals, fixing what the leader has revealed. We use this terminology only to simply exposition. The last row exhibits the ex ante probability of each event.

Similarly, we obtain the tables exhibiting μ_b under each strategy of the follower, contingent on other strategies and signals revealed by the leader. We call the leader's strategy to present the strong signal s_A when he has both t_A and s_A the "antagonistic" strategy.

	only t_B	only s_B	both s_B& t_B
str. t_B	$\mu_b = 1-\varepsilon$	$\mu_b = \dfrac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}$	$\mu_b = 1-\varepsilon$
str. s_B	$\mu_b = 1/2$	$\mu_b = \dfrac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}$	$\mu_b = \dfrac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}$
ex ante prob.	$\dfrac{\varepsilon(1-\varepsilon)}{4}$	$\dfrac{(1-\varepsilon)^2}{8}+\dfrac{\varepsilon^2}{8}$	$\dfrac{(1-\varepsilon)^2}{8}+\dfrac{\varepsilon^2}{8}$

Table 2. The leader who plays the antagonistic strategy presented t_A.

	only t_B	only s_B	both s_B & t_B
str. t_B	$\mu_b = \varepsilon$	$\mu_b = 1/2$	$\mu_b = \varepsilon$
str. s_B	$\mu_b = \dfrac{\varepsilon^2}{(1-\varepsilon)^2+\varepsilon^2}$	$\mu_b = 1/2$	$\mu_b = 1/2$
ex ante prob.	$\dfrac{(1-\varepsilon)^2}{4}+\dfrac{\varepsilon^2}{4}$	$\dfrac{\varepsilon(1-\varepsilon)}{2}$	$\dfrac{\varepsilon(1-\varepsilon)}{2}$

Table 3. The leader who plays the antagonistic strategy presented s_A.

We omit the table for the case in which the leader who plays the conciliatory strategy has presented s_A. For $\delta = 1/2$, the entries in the first two rows of this table are the same as those in Table 3 and the entries in the third row of this table is equal to those in Table 3 multiplied by 1/2.

We call a function $u : [0,1] \to R$ *concave at* 1/2 if

$$\frac{1}{2}u(\frac{1}{2}+x)+\frac{1}{2}u(\frac{1}{2}-x) \leq u(\frac{1}{2}), \text{ for every } x \in (0,1/2];$$

and *convex at* 1/2 if

$$\frac{1}{2}u(\frac{1}{2}+x)+\frac{1}{2}u(\frac{1}{2}-x) \geq u(\frac{1}{2}), \text{ for every } x \in (0,1/2].$$

Proposition 1 ***(i)*** *Suppose the leader plays the conciliatory strategy. If the follower's utility is concave at* 1/2*, he should respond with the weak signal to the weak signal. If the follower's utility is convex at* 1/2*, he should respond with the strong signal to the weak signal.*

(ii) *Suppose the leader plays the antagonistic strategy. If the follower's utility is concave on* [1/2, 1]*,*

he should respond with the weak signal to the weak signal. If the follower's utility is convex on $[1/2, 1]$*, he should respond with the strong signal to the weak signal.*
(iii) *Independent of the leader's strategy: if the follower's utility is concave on* $[0, 1/2]$*, he should respond with the weak signal to the strong signal; and if the follower's utility is convex on* $[0, 1/2]$*, he should respond with the strong signal to the strong signal.*

Proof. We provide the proof for part (i) and omit the proofs for parts (ii) and (iii) since they are similar.

The only two events in which the follower obtains different utilities under different strategies are when he has only signal t_B and when he has both signals s_B and t_B. Contingent on the union of the two events, the follower's expected utility is

$$\frac{1}{2}u_B(\varepsilon) + \frac{1}{2}u_B(1-\varepsilon),$$

when he plays the strategy of responding with the strong signal, and

$$u_B(1/2),$$

when he plays the strategy of responding with the weak signal. Since $\varepsilon/2 + (1-\varepsilon)/2 = 1/2$, the concavity (convexity) of u_B makes the former expression larger (smaller) than the latter expression.■

The proof of Proposition 1 is straightforward; nevertheless, we find it helpful to explain the argument verbally and intuitively. Notice first that the process of information revelation has the following martingale property: the audience's belief regarding the state of the world, at any point in time, is determined by the strategies of players that have already moved and the signals that have been revealed, and are independent of the strategies of the players who will move in the future. Thus, the follower's strategy does not affect the expected beliefs of the audience, and affects only the dispersion of the beliefs. This dispersion is higher when the follower plays the strategy of responding with the strong signal. To see this, note that if the follower plays this strategy, then, when he presents the strong signal the audience gets convinced (that is, attaches a high probability) that the state is b, but when he presents the weak signal, the audience infers that he lacks the strong signal, and gets convinced that the state is a. In contrast, when the follower plays the strategy of responding with the weak signal, the audience does not infer much about the state when the weak signal is presented, resulting in less dispersion in the audience's posterior.

It follows from the proof that Proposition 1 generalizes to priors over states of nature other than 1/2, where the concavity (convexity) at the prior means that the agent's utility of the prior is higher (lower) than the expected utility of the lottery induced by the signal. We conjecture that all our results generalize to other priors under this meaning of concavity and convexity.

Since our results depends on the curvature of the follower's utility function, one would like to know which properties seem to be plausible. The answer depends, of course, on the way the audience uses the information conveyed by the discussants and the subsequent actions of the audience.

Nevertheless, we conjecture that in many situations of interest, the utility functions may be convex on the interval $[0, 1/2]$, concave on the interval $[1/2, 1]$, and have a kink, so that they are concave at $1/2$. The example of the utility function $u_I(\mu_i) = (\mu_i)^3 + 3(\mu_i)^2(1 - \mu_i)$ illustrates how concavity on $[1/2, 1]$ and convexity on $[0, 1/2]$ arise naturally in some situations. This conjecture is also supported by the observation that the fifty-fifty belief is "pivotal" in certain applications, and there seems to be a bigger difference between the posterior being equal to .49 and .51 than between .01 and .03, or .97 and .99. Given the fifty-fifty prior, the discussants may have some loss aversion, analogous to that experimentally demonstrated by Kahneman and Tversky (1979). Figure 1 illustrates a utility function with these properties. (Our analysis focuses on utility functions with these properties, but can be adapted to utility functions with other curvature properties as well.)

Under these assumptions on the utility function, the follower should respond by, or try to build a reputation for, presenting weak arguments in response to weak arguments and presenting strong arguments in response to strong arguments, independent of the leader's strategy. Although the prediction that the follower may not want to reveal his strongest evidence seems surprising at first, in practice there are instances in which discussants behave in this manner. It is not uncommon that people refuse to respond to arguments against their case if they find the opponent's arguments weak or irrelevant, as illustrated by the examples in the introduction.

Of course, responding with a weak argument when having both arguments is not an equilibrium strategy without commitment. This is because even if the audience expects the follower to present the weak argument, it is still better to present the strong argument when it is available as it changes the audience's posterior favorably (and the implicit inference that the weak evidence is absent does not matter for the audience's posterior when $\delta = 1/2$). In the less extreme case where δ is lower than $1/2$, however, it can be optimal to present the weaker argument when having both even without commitment. As long as t_B is not too weak compared to s_B, presenting the stronger argument when the audience expects the weaker one is damaging because the audience will infer that the weaker evidence does not exist. In contrast, if the weaker argument is presented, the audience does not draw any inference about whether the stronger evidence exists. (See Appendix A for details.)

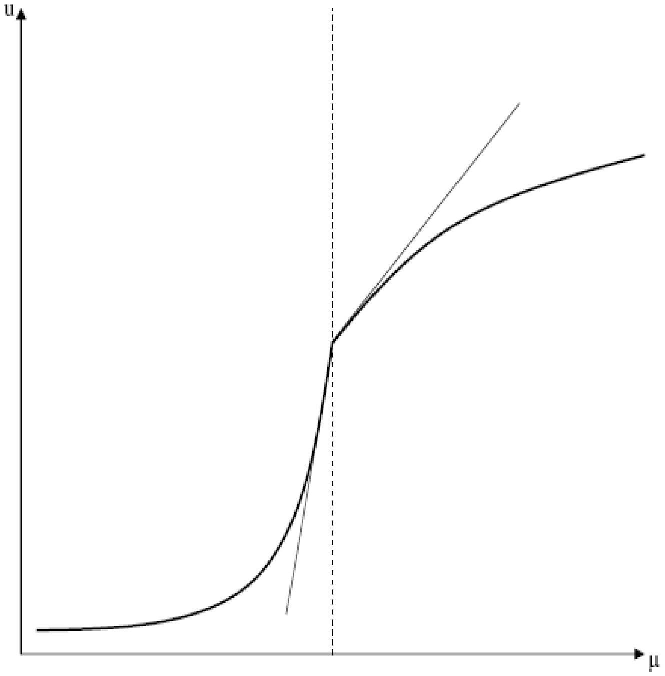

Figure 1. The utility function u, depicted in bold, has a kink at 1/2.

3.2 The Leader's Problem: Always Raise the Strongest Argument First?

We now turn to the leader's problem. What argument should the leader raise in anticipation of the follower's response? Is it always wise for the leader to raise his strongest argument first?

In the following analysis, we make the same assumptions as in section 3.1 that $\rho = 0$ and $\varepsilon < \delta = \frac{1}{2}$. We also assume that the follower's utility function is convex on $[0, 1/2]$ and concave on $[1/2, 1]$ and it is concave at 1/2. (Similar results can be derived if we make alternative assumptions on the convexity or concavity of the utility function.) According to Proposition 1, the follower responds with a weak signal to a weak signal, and responds with a strong signal to a strong signal.

Tables $1', 2'$ and $3'$ below exhibit μ_a, the audience's posterior that $\omega = a$, contingent on the strategies and signals of the leader and incorporating the best responses of the follower. They contain the relevant rows from Tables 1, 2 and 3.[5]

	only t_B	only s_B	both s_B & t_B	neither s_B nor t_B
str. t_B	$\mu_a = 1/2$	$\mu_a = \varepsilon$	$\mu_a = 1/2$	$\mu_a = 1 - \varepsilon$
ex ante prob.	1/8	1/8	1/8	1/8

Table 1′. The leader who plays the conciliatory strategy has either t_A or both t_A and s_A.

So he presents t_A and the follower responds with strategy t_B.

	only t_B	only s_B	both s_B & t_B	neither s_B nor t_B
str. t_B	$\mu_a = \varepsilon$	$\mu_a = \dfrac{\varepsilon^2}{(1-\varepsilon)^2 + \varepsilon^2}$	$\mu_a = \varepsilon$	$\mu_a = 1/2$
ex ante prob.	$\dfrac{\varepsilon(1-\varepsilon)}{4}$	$\dfrac{(1-\varepsilon)^2}{8} + \dfrac{\varepsilon^2}{8}$	$\dfrac{(1-\varepsilon)^2}{8} + \dfrac{\varepsilon^2}{8}$	$\dfrac{\varepsilon(1-\varepsilon)}{4}$

Table 2′. The leader who plays the antagonistic strategy has only t_A.

So he presents signal t_A and the follower responds with strategy t_B.

	only t_B	only s_B	both s_B & t_B	neither s_B or t_B
str. s_B	$\mu_a = \dfrac{(1-\varepsilon)^2}{(1-\varepsilon)^2 + \varepsilon^2}$	$\mu_a = 1/2$	$\mu_a = 1/2$	$\mu_a = \dfrac{(1-\varepsilon)^2}{(1-\varepsilon)^2 + \varepsilon^2}$
ex ante prob.	$\dfrac{(1-\varepsilon)^2}{8} + \dfrac{\varepsilon^2}{8}$	$\dfrac{\varepsilon(1-\varepsilon)}{4}$	$\dfrac{\varepsilon(1-\varepsilon)}{4}$	$\dfrac{(1-\varepsilon)^2}{8} + \dfrac{\varepsilon^2}{8}$

Table 3′. The leader who plays the antagonistic strategy has both s_A and t_A.

So he present s_A and the follower responds with strategy s_B.

We omit the table for the case in which the leader has only s_A and thus presents s_A (independent of whether he plays the antagonistic or conciliatory strategy). For $\delta = 1/2$, the entries are the same as those in Table 3′.

Proposition 2 *Suppose both players' utility functions are convex on* $[0, 1/2]$*, concave on* $[1/2, 1]$ *and concave at* $1/2$*. Then the leader should play the conciliatory strategy.*

Proof. Since the distributions of posteriors are the same in the event "the leader has only s_A" and the event "the leader has neither s_A nor t_A," we only need to consider the events "the leader has only t_A" and the event "the leader has both s_A and t_A."

[5] The last rows in Tables 1′, 2′, and 3′ are the same as the last rows of Table 1, 2, and 3, respectively, except in Table 3′ where it is multiplied by 1/2 because Table 3 contains also the event that the leader has only s_A.

To prove the proposition, we need to show that

$$\frac{1}{8}u_A(\varepsilon)+\frac{1}{4}u_A(1/2)+\frac{1}{8}u_A(1-\varepsilon)$$

$$\begin{aligned}\geq\quad&\left(\frac{\varepsilon(1-\varepsilon)}{4}+\frac{(1-\varepsilon)^2+\varepsilon^2}{8}\right)u_A(\varepsilon)+\frac{(1-\varepsilon)^2+\varepsilon^2}{8}u_A\left(\frac{\varepsilon^2}{(1-\varepsilon)^2+\varepsilon^2}\right)\\&+\frac{3\varepsilon(1-\varepsilon)}{4}u_A(1/2)+\frac{(1-\varepsilon)^2+\varepsilon^2}{4}u_A\left(\frac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}\right).\end{aligned}$$

Since $\frac{\varepsilon(1-\varepsilon)}{4}+\frac{(1-\varepsilon)^2+\varepsilon^2}{8}=\frac{1}{8}$, we only need to show that

$$\frac{1}{4}u_A(1/2)+\frac{1}{8}u_A(1-\varepsilon)$$

$$\geq\frac{(1-\varepsilon)^2+\varepsilon^2}{8}u_A\left(\frac{\varepsilon^2}{(1-\varepsilon)^2+\varepsilon^2}\right)+\frac{3\varepsilon(1-\varepsilon)}{4}u_A(1/2)+\frac{(1-\varepsilon)^2+\varepsilon^2}{4}u_A\left(\frac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}\right).$$

Concavity of u_A at $1/2$ implies that

$$\frac{(1-\varepsilon)^2+\varepsilon^2}{8}u_A\left(\frac{\varepsilon^2}{(1-\varepsilon)^2+\varepsilon^2}\right)+\frac{(1-\varepsilon)^2+\varepsilon^2}{8}u_A\left(\frac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}\right)\leq\frac{(1-\varepsilon)^2+\varepsilon^2}{4}u_A(1/2).$$

Since $\frac{1}{2}<(1-\varepsilon)<\frac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}$, concavity of u_A on $[1/2,1]$ implies that

$$\frac{\left((1-\varepsilon)^2+\varepsilon^2\right)}{8}u_A\left(\frac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}\right)+\frac{(2\varepsilon(1-\varepsilon))}{8}u_A(1/2)\leq\frac{1}{8}u_A(1-\varepsilon).$$

Hence

$$\begin{aligned}&\frac{(1-\varepsilon)^2+\varepsilon^2}{8}u_A\left(\frac{\varepsilon^2}{(1-\varepsilon)^2+\varepsilon^2}\right)+\frac{3\varepsilon(1-\varepsilon)}{4}u_A(1/2)+\frac{(1-\varepsilon)^2+\varepsilon^2}{4}u_A\left(\frac{(1-\varepsilon)^2}{(1-\varepsilon)^2+\varepsilon^2}\right)\\\leq\quad&\left(\frac{(1-\varepsilon)^2+\varepsilon^2}{4}+\frac{\varepsilon(1-\varepsilon)}{2}\right)u_A(1/2)+\frac{1}{8}u_A(1-\varepsilon)=\frac{1}{4}u_A(1/2)+\frac{1}{8}u_A(1-\varepsilon).\end{aligned}$$

So the leader should play the conciliatory strategy. ■

To gain some intuition, note that there are two events in which the leader's strategy matters: (i) the leader only has the weak signal, and (ii) the leader has both the weak and the strong signals.

If the leader plays the conciliatory strategy, then in either one of these two events, he presents the weak signal and the follower responds with the strategy of presenting the weak signal when he has both. As such, $\mu_a=1/2$ if the follower presents a weak signal, $\mu_a=\varepsilon$ if the follower presents a strong signal, and $\mu_a=(1-\varepsilon)$ if the follower presents neither.

If the leader plays the antagonistic strategy, then the follower's strategy depends on whether the leader reveals the strong signal or the weak signal. Specifically, if the leader reveals the strong signal, then the follower responds with the strong signal when he has both. If the follower presents a strong signal, then $\mu_a = 1/2$. If the follower fails to present a strong signal, then μ_a is above $1/2$; in fact, μ_a in this case is higher than $(1-\varepsilon)$, because there are now two informative signals in favor of state a. If the leader reveals the weak signal, then the follower responds with the weak signal when he has both. The posterior μ_a is equal to ε if the follower presents a weak signal, is lower than ε if the follower presents a strong signal, and is equal to $1/2$ if the follower presents neither signal.

To summarize, if the leader plays the antagonistic strategy, the posteriors induced have more dispersion on $[1/2, 1]$ and also around $1/2$ than if he plays the conciliatory strategy. So if the leader's utility function is concave on $[1/2, 1]$ and concave at $1/2$, he should play the conciliatory strategy.

3.3 Mixed Strategies

One natural question is what happens if the discussants can commit to mixed strategies. This is interesting because committing to revealing arguments in a random order might have strategic advantages. Our main results generalize to mixed strategies under the Kahneman-Tversky preference.

To see this, note that one implication of Proposition 1 is that under the Kahneman-Tversky preference, the follower's best reply is to respond with the strong signal to the strong signal and with the weak signal to the weak signal, *independent of the leader's strategy.* Given this dominance, even when mixed strategies are allowed, the follower's optimal strategy does not change, and therefore the leader's optimal strategy does not change either. To summarize, we have the following proposition.

Proposition 3 *Suppose both players' utility functions are convex on* $[0, 1/2]$*, concave on* $[1/2, 1]$*, and concave at* $1/2$*. Even when mixed strategies are allowed, the follower should respond with the weak signal to the weak signal and with the strong signal to the strong signal, and the leader should play the conciliatory strategy.*

4 Correlated Versus Independent Evidence

4.1 The Follower's Problem: Respond Directly to the Opponent's Argument or Change Topic?

To illustrate the question we address in this section, recall the following problem discussed in Glazer and Rubinstein (2001): Suppose you are trying to convince an audience that in most capital cities, the level of education has risen recently. Your opponent brings hard evidence showing that the level of education in Bangkok has fallen. Should you respond by bringing similar evidence showing that the level of education has risen from Manila or similar evidence from Mexico City (or perhaps other more distant cities)?

Glazer and Rubinstein present evidence from questionnaires to argue that most people would recommend bringing the evidence from Manila, which seems more similar to Bangkok. Similarly, most people would recommend bringing the evidence from Brussels to counter the evidence from Amsterdam. Glazer and Rubinstein also argue that this phenomenon is not confined to cases in which people have implicit beliefs about some correlation between arguments.

Although we believe that correlation between arguments plays an important role in debates, our results show that an explanation for Glazer and Rubinstein's experiment based on correlation patterns is not as straightforward as one might expect, and it requires specific assumptions on agents' utilities which are violated, for example, in the voting scenario described in section 2.

Suppose that $\rho \neq 0$, and $\delta = \varepsilon < 1/2$. That is, arguments s_A and $\neg s_B$ are conditionally correlated, but any single argument is equally informative about the state of nature.

To illustrate how the audience's posterior depends on the discussants' strategies and the evidence presented, suppose first that the leader's strategy is to present signal s_A when he has both s_A and t_A (as in the case of weak versus strong evidence, we call this the "antagonistic" strategy) and he brings evidence s_A in favor of his claim. The following table exhibits the posterior belief of the audience, μ_b, under each strategy of the follower, given the signals at the follower's disposal. (Again, we omit the event that the follower has no evidence because the follower's strategy does not affect the posterior in this case.)

	only s_B	only t_B	both s_B & t_B
str. s_B	$\mu_b = 1/2$	$\mu_b = \dfrac{\varepsilon + \rho(1-\varepsilon)}{1+\rho}$	$\mu_b = 1/2$
str. t_B	$\mu_b = \varepsilon$	$\mu_b = 1/2$	$\mu_b = 1/2$
ex ante prob.	$\frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)(1+\rho)$	$\frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)$

Table 4. The leader who plays the antagonistic strategy presented s_A.

Similarly, we obtain tables exhibiting the posteriors of the audience μ_b for other strategies and signals of the discussants. We call the leader's strategy of presenting t_A when he has both s_A and t_A the "conciliatory" strategy.

Let

$$\begin{aligned}
\pi &= \varepsilon^2(1-\varepsilon)^2 + \frac{1}{2}\rho\varepsilon(1-\varepsilon)\left[\varepsilon^2 + (1-\varepsilon)^2\right], \\
\pi' &= \frac{1}{2}\varepsilon(1-\varepsilon)[\varepsilon^2 + (1-\varepsilon)^2] + \rho\varepsilon^2(1-\varepsilon)^2, \\
\pi'' &= \frac{1}{2}[\varepsilon^2 + (1-\varepsilon)^2](1-\rho)\varepsilon(1-\varepsilon),
\end{aligned}$$

and

$$\mu_1 = \frac{1-\varepsilon+\rho\varepsilon}{1+\rho},$$

$$\mu_2 = \frac{\varepsilon(1-\varepsilon)+\rho\varepsilon^2}{2\varepsilon(1-\varepsilon)+\rho[(1-\varepsilon)^2+\varepsilon^2]}.$$

	only s_B	only t_B	both s_B & t_B
str. s_B	$\mu_b = 1-\varepsilon$	$\mu_b = 1-\mu_2$	$\mu_b = 1-\varepsilon$
str. t_B	$\mu_b = 1/2$	$\mu_b = 1-\varepsilon$	$\mu_b = 1-\varepsilon$
ex ante prob.	$(1-\rho)\varepsilon^2(1-\varepsilon)^2$	π	π''

Table 5. The leader who plays the conciliatory strategy presented s_A.

	only s_B	only t_B	both s_B & t_B
str. s_B	$\mu_b = 1/2$	$\mu_b = \varepsilon$	$\mu_b = 1/2$
str. t_B	$\mu_b = \varepsilon$	$\mu_b = 1/2$	$\mu_b = 1/2$
ex ante prob.	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)$

Table 6. The leader who plays the conciliatory strategy presented t_A.

	only s_B	only t_B	both s_B & t_B
str. s_B	$\mu_b = \mu_1$	$\mu_b = 1/2$	$\mu_b = \mu_1$
str. t_B	$\mu_b = \mu_2$	$\mu_b = 1 - \varepsilon$	$\mu_b = 1 - \varepsilon$
ex ante prob.	π	$(1-\rho)\varepsilon^2(1-\varepsilon)^2$	π'

Table 7. The leader who plays the antagonistic strategy presented t_A.

We have the following result which describes the follower's best responses.

Proposition 4 ***(i)*** *Suppose the leader plays the antagonistic strategy and presents signal s_A. Then, the follower should respond with signal s_B when having both signals if his utility function is concave on $[0, 1/2]$ and should respond with signal t_B when having both signals if his utility function is convex on $[0, 1/2]$.*
(ii) *Suppose the leader plays the conciliatory strategy and presents signal s_A. Then, the follower should respond with signal s_B when having both signals if his utility function is concave on $[1/2, 1]$ and should respond with signal t_B when having both signals if his utility function is convex on $[1/2, 1]$.*
(iii) *Suppose the leader plays the conciliatory strategy and presents signal t_A. Then, the follower is indifferent between the two possible strategies.*
(iv) *Suppose the leader plays the antagonistic strategy and presents signal t_A. Then, the follower should respond with signal s_B when having both signals if his utility function is concave on $[1/2, 1]$ and concave at $1/2$ and should respond with signal t_B when having both signals if his utility function is convex on $[1/2, 1]$ and convex at $1/2$.*

Proof. By the martingale property of the process of information revelation, or by direct computation, the expected posterior beliefs are equal under the two strategies of the follower,

$$E^{\text{str. } s_B}(\mu_b) = E^{\text{str. } t_B}(\mu_b). \tag{1}$$

Part (i) follows since in Table 4,

$$\frac{\varepsilon + \rho(1-\varepsilon)}{1+\rho} \in (\varepsilon, 1/2).$$

In order to obtain part (ii), notice that in Table 5,

$$1 - \mu_2 = \frac{\varepsilon(1-\varepsilon) + \rho(1-\varepsilon)^2}{2\varepsilon(1-\varepsilon) + \rho\left[\varepsilon^2 + (1-\varepsilon)^2\right]} \in (1/2, 1-\varepsilon).$$

This together with equation (1) yields (ii).

Part (iii) follows immediately from Table 6.

In order to obtain part (iv), notice that

$$\mu_1 = \frac{1-\varepsilon+\rho\varepsilon}{1+\rho} \in (1/2, 1-\varepsilon)\,,$$

$$\mu_2 = \frac{\varepsilon(1-\varepsilon)+\rho\varepsilon^2}{2\varepsilon(1-\varepsilon)+\rho\left[\varepsilon^2+(1-\varepsilon)^2\right]} < 1/2.$$

So, we compare two lotteries.[6] The lottery generated by strategy s_B has outcomes $1/2$ and μ_1; and the lottery generated by strategy t_B has outcomes μ_2 and $1-\varepsilon$, where

$$\mu_2 < 1/2 < \mu_1 < 1-\varepsilon.$$

Next, we use Figure 2 to illustrate a graphical argument. By concavity of the utility function on $[1/2, 1]$, the line passing through $(1/2, u_B(1/2))$ and $(\mu_1, u_B(\mu_1))$ is steeper than the line passing through $(1/2, u_B(1/2))$ and $(\mu_3, u_B(1-\varepsilon)$. By concavity at $1/2$, the line passing through $(\mu_2, u_B(\mu_2))$ and $(1/2, u_B(1/2))$ is steeper than the line passing through $(1/2, u_B(1/2))$ and $(1-\mu_2, u_B(1-\mu_2))$.

Recall that $1-\mu_2 < 1-\varepsilon$. Thus, by concavity on $[1/2, 1]$, the line passing through $(\mu_2, u_B(\mu_2))$ and $(1/2, u_B(1/2))$ is steeper than the line passing through $(1/2, u_B(1/2))$ and $(1-\varepsilon, u_B(1-\varepsilon))$.

If $(\mu_2, u_B(\mu_2))$ belonged to the line passing through $(1/2, u_B(1/2))$ and $(1-\varepsilon, u_B(1-\varepsilon))$, then $(E(\mu),\ E^{\text{str. } t_B}\ u_B(\mu))$ would belong to that line as well. However, since the line passing through $(\mu_2, u_B(\mu_2))$ and $(1/2, u_B(1/2))$ is steeper than the line passing though $(1/2, u_B(1/2))$ and $(1-\varepsilon, u_B(1-\varepsilon))$, it must be the case that $(E(\mu_b),\ E^{\text{str. } t_B}\ u_B(\mu_b))$ lies below the line passing through $(1/2, u_B(1/2))$ and $(1-\varepsilon, u_B(1-\varepsilon))$. On the other hand, $(E(\mu_b),\ E^{\text{str. } s_B}\ u_B(\mu_b))$ lies above the line passing through $(1/2, u_B(1/2))$ and $(1-\varepsilon, u_B(1-\varepsilon))$, because it belongs to the steeper line passing through $(1/2, u_B(1/2))$ and $(\mu_1, u_B(\mu_1))$. This yields that

$$E^{\text{str. } t_B}\ u_B(\mu_b) \leq E^{\text{str. } s_B}\ u_B(\mu_b).$$

■

[6] Strictly speaking these are not lotteries because the ex ante probabilities do not add up to 1, but a simple normalization transforms these into lotteries and the argument is not affected. For expositional convenience, we call them lotteries here and in the proofs of other propositions as well.

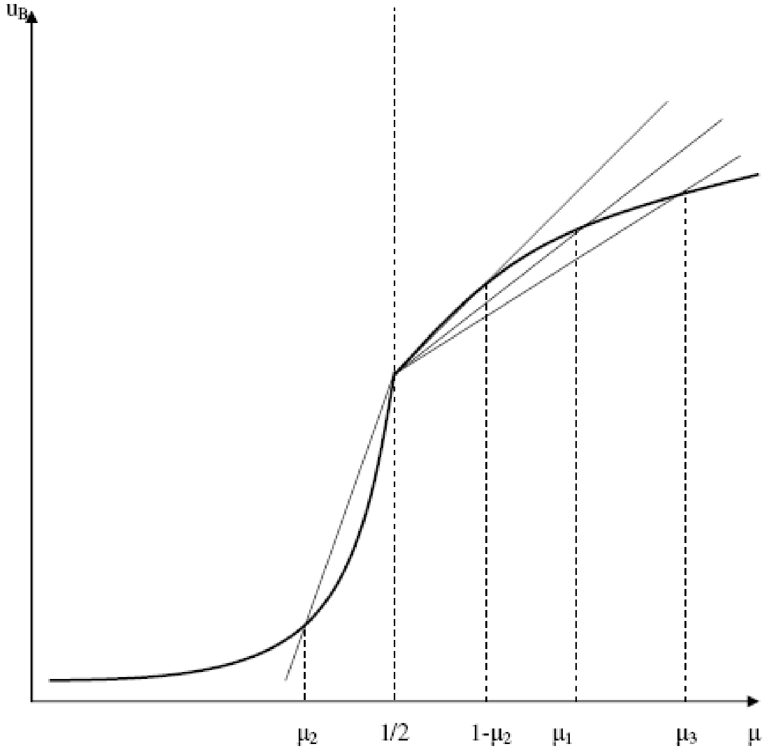

Figure 2. The line passing through $(\mu_2, u_B(\mu_2))$ and $(1/2, u_B(1/2))$ is steeper than the line passing through $(1/2, u_B(1/2))$ and $(1-\mu_2, u_B(1-\mu_2))$, which in turn is steeper than the line passing through $(1/2, u_B(1/2))$ and $(1-\varepsilon, u_B(1-\varepsilon))$. Also, the line passing through $(1/2, u_B(1/2))$ and $(\mu_1, u_B(\mu_1))$ is steeper than the line passing through $(1/2, u_B(1/2))$ and $(1-\varepsilon, u_B(1-\varepsilon))$.

To understand the intuition behind Proposition 4, note that because of the correlation between certain signals, what strategy is more informative (induces more dispersion in the audience's posterior) for the follower is now determined endogenously by the leader's strategy as well as the evidence (or the lack thereof) revealed by him. Specifically, in part (i), when the leader who plays the antagonistic strategy has presented signal s_A, the audience's posterior μ_b can be no higher than 1/2. If the follower plays strategy s_B (t_B) and presents the signal s_B (t_B), then μ_b is equal to 1/2. But if the follower plays strategy s_B (t_B) but presents the signal t_B (s_B), then μ_b is lower than 1/2. Because s_A and $\neg s_B$ are positively correlated, the revelation of s_A by the leader already indicates that it is unlikely for the follower to have evidence s_B. As such, the audience's posterior μ_b is higher if the follower plays strategy s_B but presents the signal t_B (revealing that he does not have s_B) than if the follower plays strategy t_B but presents the signal s_B (revealing that he does not have t_B). So, there is more dispersion in the audience's belief if the follower plays the strategy t_B than if he plays strategy s_B. Hence he is better off playing strategy t_B if his utility function is convex on $[0, 1/2]$ and he is better off playing strategy s_B if his utility function is concave on $[0, 1/2]$.

The intuition for part (ii) is similar to that for part (i): if the leader plays the conciliatory strategy but presents signal s_A, then the audience's posterior μ_b is at least $1/2$ as long as the follower shows some evidence in favor of his claim. If the follower's strategy is t_B, the posterior μ_b is above $1/2$ if he presents t_B and it is equal to $1/2$ if he presents s_B. If the follower's strategy is s_B, the posterior μ_b is above $1/2$ if he presents s_B and it is still above $1/2$ even when he presents t_B. Again, this is because s_A and $\neg s_B$ are positively correlated. Since the leader already reveals that he has evidence s_A but no t_A, even if the follower reveals that he only has evidence t_B and does not have evidence s_B, the audience's posterior is still above $1/2$. In short, because of the correlation, the follower's strategy s_B induces a lower dispersion in the audience's belief, and hence it is the optimal strategy if his utility function is concave on $[1/2, 1]$, while the strategy t_B is optimal if his utility function is convex on $[1/2, 1]$.

Part (iii) is straightforward: if the leader plays the conciliatory strategy and he presents t_A, then the distribution of posteriors is the same no matter what strategy the follower plays.

Part (iv) describes the optimal strategy for the follower if the leader plays the antagonistic strategy but presents signal t_A, that is, the leader reveals that he has evidence t_A, but not s_A. If the follower responds with strategy s_B, then the posterior μ_b is higher than $1/2$ if he presents s_B and is equal to $1/2$ if he presents t_B. If the follower responds with strategy t_B, then the posterior μ_b is higher than $1/2$ if he presents t_B, but it is lower than $1/2$ if he presents s_B. Because of the positive correlation between $\neg s_A$ and s_B, the revelation that the follower has s_B but no t_B does not sway the posterior favorably enough for it to exceed $1/2$. So if the follower's utility exhibits loss aversion at $1/2$ and concavity on $[1/2, 1]$, then he should play the strategy s_B. Similarly, if the follower's utility is convex at $1/2$ and convex on $[1/2, 1]$, then he should play the strategy t_B.

Although in deriving the results we have assumed commitment to strategies, we would like to point out that the optimal strategies we have identified for the follower are equilibrium strategies even without commitment. For example, if the audience expects that the follower will respond to the leader's signal s_A with strategy t_B, then indeed the follower should optimally choose the strategy

t_B because with this expectation, the audience believes that the follower does not have signal t_B when s_B is revealed but believes that the follower may also have s_B when t_B is revealed.

4.2 The Leader's Problem

Let

$$\begin{aligned} \mu_3 &= \frac{(1-\varepsilon)^2 + \rho\varepsilon(1-\varepsilon)}{(1-\varepsilon)^2 + 2\rho\varepsilon(1-\varepsilon) + \varepsilon^2}, \\ \mu_4 &= \frac{\left[(1-\varepsilon)^2 + \rho\varepsilon(1-\varepsilon)\right](1-\varepsilon)}{(1-\varepsilon)^3 + \varepsilon^3 + \rho\varepsilon(1-\varepsilon)}. \end{aligned}$$

In Table 8 and Table 9, we summarize the audience's posterior μ_a induced by the leader's strategies by incorporating the follower's best responses and also the corresponding ex ante probabilities. The variables μ_1, μ_2, π, π' and π'' have been defined in the previous section and details of how we derive these posteriors are in Appendix C.

μ_a	$1-\mu_1$	$1/2$	$1-\varepsilon$	$\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}$	μ_4
ex ante prob.	$\pi+\pi'$	$(1-\varepsilon)\varepsilon+$ $(1-\rho)\varepsilon^2(1-\varepsilon)^2$	$\frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)$	π''	$\frac{1}{2}(1-\varepsilon)^3+\frac{1}{2}\varepsilon^3$ $+\frac{1}{2}\rho\varepsilon(1-\varepsilon)$

Table 8. The leader plays the antagonistic strategy.

μ_a	ε	μ_2	$1/2$	$1-\varepsilon$	μ_3	$\frac{(1-\varepsilon)^3}{(1-\varepsilon)^3+\varepsilon^3}$
ex ante prob.	$(1-\rho)\varepsilon^2(1-\varepsilon)^2+\pi''$	π	$\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	π'	$\frac{1}{2}[(1-\varepsilon)^3+\varepsilon^3]$

Table 9. The leader plays the conciliatory strategy.

We call a function $u : [0,1] \to R$ *strictly concave at* $1/2$ if

$$\frac{1}{2}u\left(\frac{1}{2}+x\right)+\frac{1}{2}u\left(\frac{1}{2}-x\right) < u\left(\frac{1}{2}\right), \text{ for every } x \in (0,1/2];$$

Proposition 5 *Suppose both players' utility functions are continuous on* $[0,1]$*, convex on* $[0,1/2]$*, concave on* $[1/2,1]$*, and strictly concave at* $1/2$*. Then the leader should play the antagonistic strategy if* ρ *is sufficiently high.*

Proof. Decompose the lottery induced by each strategy into two: one corresponding to the entries of the last two columns of Tables 8 and 9, and the other corresponding to the entries of the first three columns. Consider first the lotteries corresponding to the entries of the last two columns. Since

$$\pi'' \cdot \frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2} + \frac{1}{2}\left[(1-\varepsilon)^3+\varepsilon^3+\rho\varepsilon(1-\varepsilon)\right]\mu_4 = \pi'\mu_3 + \frac{1}{2}\left[(1-\varepsilon)^3+\varepsilon^3\right]\cdot\frac{(1-\varepsilon)^3}{(1-\varepsilon)^3+\varepsilon^3},$$

these lotteries have the same mean.

Since $\frac{1}{2} < \mu_3 < \frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2} < \mu_4 < \frac{(1-\varepsilon)^3}{(1-\varepsilon)^3+\varepsilon^3}$ for $\rho \in (0,1)$, it follows that if u_A is concave on $[1/2, 1]$, then

$$\begin{aligned} & \pi'' u_A\left(\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}\right) + \frac{1}{2}\left[(1-\varepsilon)^3+\varepsilon^3+\rho\varepsilon(1-\varepsilon)\right]u_A(\mu_4) \\ \geq\ & \pi' u_A(\mu_3) + \frac{(1-\varepsilon)^3}{(1-\varepsilon)^3+\varepsilon^3}\cdot u_A\left(\frac{1}{2}\left((1-\varepsilon)^3+\varepsilon^3\right)\right). \end{aligned} \tag{2}$$

That is, for the last two columns of Tables 8 and 9, the leader prefers the lottery induced by strategy s_A than that induced by strategy t_A if u_A is concave on $[1/2, 1]$.

The other columns of Tables 8 and 9 correspond to lotteries with mean 1/2, since

$$(\pi+\pi')(1-\mu_1) + \left[\frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)\right](1-\varepsilon) = \left[\pi+\pi'+\frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)\right]\frac{1}{2},$$

and

$$\left[(1-\rho)\varepsilon^2(1-\varepsilon)^2+\pi''\right]\varepsilon + \pi\mu_2 + \frac{1}{2}\varepsilon(1-\varepsilon)(1-\varepsilon) = \left[(1-\rho)\varepsilon^2(1-\varepsilon)^2+\pi''+\pi+\frac{1}{2}\varepsilon(1-\varepsilon)\right]\frac{1}{2}.$$

So, whether the leader prefers the antagonistic strategy or the conciliatory strategy depends also on his preference over the two lotteries with mean 1/2. Let $E^A(u_A)$ and $E^C(u_A)$ denote the leader's expected payoff if he plays the antagonistic strategy and if he plays the conciliatory strategy, respectively. From inequality (2), we have

$$\begin{aligned} E^A(u_A) - E^C(u_A) \geq\ & (\pi+\pi')\, u_A(1-\mu_1) + \left[(1-\varepsilon)\varepsilon + (1-\rho)\varepsilon^2(1-\varepsilon)^2\right]u_A(1/2) \\ & + \frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)\, u_A(1-\varepsilon) - \left[(1-\rho)\varepsilon^2(1-\varepsilon)^2+\pi''\right]u_A(\varepsilon) \\ & - \pi u_A(\mu_2) - \varepsilon(1-\varepsilon)\, u_A(1/2) - \frac{1}{2}\varepsilon(1-\varepsilon)\, u_A(1-\varepsilon) \end{aligned}$$

Since $\lim_{\rho\to1}\mu_1 = \frac{1}{2}$, $\lim_{\rho\to1}\mu_2 = \varepsilon$, $\lim_{\rho\to1}\pi = \frac{1}{2}\varepsilon(1-\varepsilon)$, $\lim_{\rho\to1}\pi' = \frac{1}{2}\varepsilon(1-\varepsilon)$, $\lim_{\rho\to1}\pi'' = 0$, u_A is continuous in μ, and π, π', π'' are continuous in ρ, we have

$$\lim_{\rho\to1}\left(E^A(u_A) - E^C(u_A)\right) \geq$$

$$\varepsilon\left(1-\varepsilon\right)u_A\left(1/2\right)-\frac{1}{2}\varepsilon\left(1-\varepsilon\right)u_A\left(\varepsilon\right)-\frac{1}{2}\varepsilon\left(1-\varepsilon\right)u_A\left(\left(1-\varepsilon\right)\right).$$

If u_A is strictly concave at $1/2$, then

$$\varepsilon\left(1-\varepsilon\right)u_A\left(1/2\right)-\frac{1}{2}\varepsilon\left(1-\varepsilon\right)u_A\left(\varepsilon\right)-\frac{1}{2}\varepsilon\left(1-\varepsilon\right)u_A\left(\left(1-\varepsilon\right)\right)>0.$$

Given the continuity of the leader's payoff in ρ, the leader should play the antagonistic strategy when ρ is sufficiently close to 1. ■

The proof shows that the leader's preference over the two strategies depends on the concavity or convexity of his utility function on $[1/2, 1]$, as well as his preference over the lotteries from the first three columns of Table 8 and the first four columns of Table 9, which have mean $1/2$. However, it can be shown that his preference over the two lotteries with mean $1/2$ is in general not determined by the concavity or convexity of the leader's utility function at $1/2$, and we are only able to establish the leader's preference over these two lotteries when the correlation between s_A and $\neg s_B$ is sufficiently high.

To gain some intuition for this result, let us take a closer look at the two pairs of lotteries in turn. First, compare the lottery $\left(\mu_3, \frac{(1-\varepsilon)^3}{(1-\varepsilon)^3+\varepsilon^3}\right)$ induced by the conciliatory strategy and the lottery $\left(\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}, \mu_4\right)$ induced by the antagonistic strategy. The one induced by the conciliatory strategy has more dispersion because (conditionally) independent evidence in favor of the same state sways the audience belief more than (conditionally and positively) correlated evidence. Because the conciliatory strategy gives precedence to the independent evidence and the antagonistic strategy gives precedence to the correlated evidence, the conciliatory strategy induces more extreme posteriors than the antagonistic strategy when the leader has favorable evidence and the follower has no evidence in his favor at all. So if the leader's payoff is concave on $[1/2, 1]$, he prefers the lottery induced by the antagonistic strategy.

Next, compare the lotteries that have mean $1/2$. It is perhaps easiest to understand the result when $\rho = 1$, that is, s_A and $\neg s_B$ are perfectly correlated. In this case, the lottery induced by the antagonistic strategy is degenerate, containing only posterior equal to $1/2$ whereas the lottery induced by the conciliatory strategy contains posteriors both below and above $1/2$. If the leader's preference exhibits loss aversion, the leader prefers the lottery induced by strategy s_A. By continuity the result holds when s_A and s_B are sufficiently correlated.

In Proposition 5 we assumed strict concavity at $1/2$, but a similar result holds if we instead assume strict concavity on $[1/2, 1]$. The following example illustrates this.

Example 1: Suppose the payoff function of player A is $u_A\left(\mu_A\right) = \mu_a^3 + 3\mu_a^2\left(1-\mu_a\right)$ and the payoff function of player B is $u_B(\mu_B) = \mu_b^3 + 3\mu_b^2(1-\mu_b)$. So both players' payoff functions are strictly convex on $[0, 1/2]$ and strictly concave on $[1/2, 1]$.

Simple calculation shows that $\lim_{\rho\to 1}\mu_1 = \frac{1}{2}$, $\lim_{\rho\to 1}\mu_2 = \varepsilon$,

$$\lim_{\rho\to 1}\mu_3 = \frac{(1-\varepsilon)^2 + \varepsilon(1-\varepsilon)}{(1-\varepsilon)^2 + 2\varepsilon(1-\varepsilon) + \varepsilon^2} = 1 - \varepsilon,$$

$$\lim_{\rho\to 1}\mu_4 = \frac{\Big((1-\varepsilon)^2 + \varepsilon(1-\varepsilon)\Big)(1-\varepsilon)}{\Big((1-\varepsilon)^2 + \varepsilon(1-\varepsilon)\Big)(1-\varepsilon) + \left(\varepsilon^2 + \varepsilon(1-\varepsilon)\right)\varepsilon} = \frac{(1-\varepsilon)^2}{(1-\varepsilon)^2 + \varepsilon^2},$$

$\lim_{\rho\to 1}\pi = \frac{1}{2}\varepsilon(1-\varepsilon)$, $\lim_{\rho\to 1}\pi' = \frac{1}{2}\varepsilon(1-\varepsilon)$, $\lim_{\rho\to 1}\pi'' = 0$. So as ρ goes to 1, the leader's expected payoff when playing the antagonistic strategy is

$$\begin{aligned}\lim_{\rho\to 1} E^A(u_A) &= 2\varepsilon(1-\varepsilon)\, u_A(1/2) + \left(\varepsilon^2 - \varepsilon + 1/2\right) u_A\left(\frac{(1-\varepsilon)^2}{(1-\varepsilon)^2 + \varepsilon^2}\right) \\ &= \varepsilon(1-\varepsilon) + \left(\varepsilon^2 - \varepsilon + 1/2\right)\left(2\varepsilon^2 - 2\varepsilon + 1\right)^{-3}(\varepsilon - 1)^4\left(4\varepsilon^2 - 2\varepsilon + 1\right).\end{aligned}$$

And the leader's expected payoff when playing the conciliatory strategy is

$$\lim_{\rho\to 1} E^C(u_A) = \frac{\varepsilon(1-\varepsilon)}{2} u_A(\varepsilon) + \varepsilon(1-\varepsilon)\, u_A(1/2) + \varepsilon(1-\varepsilon)\, u_A(1-\varepsilon) + \frac{(1-\varepsilon)^3 + \varepsilon^3}{2} u_A\left(\frac{(1-\varepsilon)^3}{(1-\varepsilon)^3 + \varepsilon^3}\right)$$

$$= \frac{1}{2}\varepsilon(1-\varepsilon)\left(2\varepsilon^3 - 3\varepsilon^2 + 3\right) + \frac{1}{2}\left(3\varepsilon^2 - 3\varepsilon + 1\right)^{-2}(\varepsilon - 1)^6\left(3\varepsilon^2 - 3\varepsilon + 2\varepsilon^3 + 1\right)$$

The difference in the leader's expected payoff between the antagonistic strategy and the conciliatory strategy is

$$\lim_{\rho\to 1}\left(E^A(u_A) - E^C(u_A)\right) = \frac{(1-\varepsilon)^3(1-2\varepsilon)^3\left(\varepsilon^2 - \varepsilon + 1\right)\left(8\varepsilon^2 - 8\varepsilon + 3\right)\varepsilon^3}{2\left(2\varepsilon^2 - 2\varepsilon + 1\right)^2\left(3\varepsilon^2 - 3\varepsilon + 1\right)^2}.$$

Since $\varepsilon < 1/2$, $\varepsilon^2 - \varepsilon + 1 = (\varepsilon - 1/2)^2 + 3/4 > 0$, $8\varepsilon^2 - 8\varepsilon + 3 = 8(\varepsilon - 1/2)^2 + 1 > 0$, we have $\lim_{\rho\to 1}\left(E^A(u_A) - E^C(u_A)\right) > 0$. So if the correlation between the two pieces of evidence is sufficiently high, the leader should play the antagonistic strategy. ■

Again, we would like to point out that in addition to being the leader's optimal strategy when he can commit, the antagonistic strategy is also an equilibrium strategy even without commitment: if the audience expects the leader to give precedence to the correlated evidence, then indeed it is optimal for the leader to do so.

Returning to Glazer and Rubinstein's (2001) experiment, it follows from Proposition 4 (i) and Proposition 5 that the leader should give precedence to the correlated signal s_A and the follower should respond (or try to build the reputation for responding) with the independent signal t_B to the correlated signal s_A, opposite of what the experimental evidence from Glazer and Rubinstein (2001) suggests, at least when agents have preferences which are concave above the prior, convex below the prior, exhibit some form of loss aversion, and the correlation coefficient is high.

In our model, the audience makes inference from the discussant's strategy as well as the evidence presented. It seems more consistent with Glazer and Rubinstein's experiment, however, that the audience does not make any inference from the leader's strategy about what other evidence the leader may or may not have. Our results still apply to this setting though. It follows from Proposition 4 (i) that the follower's best response to the antagonistic strategy is to play "independent to correlated" if his utility function is concave on $[0, 1/2]$ and to play "correlated to correlated" if his utility function is convex on $[0, 1/2]$. Thus, the experimental evidence from Glazer and Rubinstein can be explained by the correlation between the education levels of geographically or culturally similar cities, and the agents' risk aversion with respect to the audience's posterior belief. This conclusion is confirmed by the costly information acquisition model in Appendix B.

Although we have restrict attention to pure strategies for simplicity, the results extend to mixed strategies as well. Specifically, we can show that under the Kahneman and Tversky preference, the antagonistic strategy of the leader dominates the conciliatory strategy for ρ sufficiently close to 1. The follower's best response to the antagonistic strategy is to respond with the independent signal to the correlated signal, and to respond with the correlated signal to the independent signal. It follows that we have the same equilibrium even if mixed strategies are allowed. (We omit the details here because the analysis is routine but tedious.)

5 Discussion

5.1 Equilibria without Commitment

In this subsection, we discuss Perfect Bayesian Equilbria of the game in which agents cannot commit to strategies. Without commitment, equilibrium requires that the agents' strategies are optimal even after they learn their signals.

We show that this game typically has multiple equilibria since equilibrium conditions impose weak constraints on the strategies of both discussants. The weak versus strong case with $0 < \varepsilon < \delta = 1/2$ and $\rho = 0$ is an exception. In this case, both discussants have an incentive to reveal the strong signal when he has it, independent of the audience's belief regarding their strategies. Accordingly, when $0 < \varepsilon < \delta = 1/2$ and $\rho = 0$, the game has a unique equilibrium in which each discussant plays an antagonistic strategy of presenting the strong signal whenever it is available.

Now consider the case in which $\delta < 1/2$ and ε and δ are not too different (for example, $\varepsilon = 1/4$ and $\delta = 1/3$). To demonstrate the multiplicity of equilibria in the simplest way, suppose that there is only one discussant. Suppose also that the audience believes that the discussant's strategy is to reveal the weaker signal if he has both signals. We show below that it is optimal for the discussant to play this strategy.

Under the audience's belief about the discussant's strategy, if he reveals the weaker signal, the audience updates its belief such that the probability of the state being the one preferred by the discussant is $1 - \delta$. If he reveals the stronger signal, however, the audience infers that the discussant does not have the weaker signal, and updates its belief such that the probability of the state being the one preferred by the discussant is

$$\frac{(1-\varepsilon)\delta}{(1-\varepsilon)\delta + (1-\delta)\varepsilon}.$$

This expression is below $(1 - \delta)$ if $(1 - \delta)^2\varepsilon > \delta^2(1 - \varepsilon)$, which holds when ε and δ are close and lower than 1/2.

Similarly, if the audience believes that the discussant's strategy is to reveal the stronger signal when he has both signals, then the discussant indeed has an incentive to play this strategy, for any values of ε and δ. Thus, when both the strong and weak signals are informative and ε and δ are sufficiently close, we have two equilibria with different outcomes.

A similar problem of multiplicity of equilibria arises in the case of correlated versus independent evidence. Suppose for simplicity that $\rho = 1$ so that a pair of signals has perfect (negatively) correlation.[7] Consider the follower first. The only situation in which his strategy matters is when the leader plays the conciliatory strategy and reveals the independent signal because otherwise the leader has already revealed whether the follower has the correlated signal.

Suppose first that the audience believes that the follower will respond with the independent signal when having both signals. If the follower presents the independent signal, this is the only

[7] This case requires a caveat since we assume $\rho = 1$. Specifically, suppose that the audience believes that the leader play the antagonistic strategy, but the leader deviates to the conciliatory strategy and reveals the independent signal when having both. Suppose further that the follower does not have the independent signal, and therefore reveals no signal. This is a probability zero event given the audience's belief about strategies. We assume that in the equilibrium described in the text, the audience believes in this event that it is at least as likely that the follower has the correlated signal in favor of his preferred state as that the leader has the correlated signal in favor of his preferred state.

information he reveals. If the follower presents the correlated signal, however, he also reveals that he does not have the independent signal. The follower clearly prefers the former to the latter, and this is confirm the audience's belief that the follower responds with the independent signal. A similar argument applies if the audience believes that the follower will respond with the correlated signal when having both – again, the belief is self-fulfilling.

Consider now the leader. Suppose that the audience believes that the leader plays the conciliatory strategy, and that the follower respond with an independent signal (if he has both) to an independent signal. Suppose the leader has both signals. If the leader reveals the independent signal, the audience learn about two independent signals or learns about the independent signal of the leader and the lack of any signal of the follower, depending on whether the follower has the independent signal. If the leader reveals the correlated signal, then the audience believes that he does not have the independent signal, and possibly learns about the independent signal of the follower (if he has it). The leader clearly prefers the conciliatory strategy. A similar analysis applies if the audience believes that the leader plays the antagonistic strategy, and that the follower responds to an independent signal with an independent signal. Because beliefs about the agents' strategies are self-fulfilling in the game without commitment, multiple equilibria arise.

5.2 Motivation for the Commitment Assumption

The game without commitment seems an appropriate model if one wants to provide some advice on how to argue in a single and isolated debate. Unfortunately, as shown in the previous section, this model typically has multiple equilibria, implying that many ways of argumentation are consistent with equilibrium. Thus, we are unable to provide any advice how how to argue in a single debate.

Instead, we study the game with commitment in this paper, motivated largely by reputation effects. The starting point is a somewhat different, but still important question. Suppose that agents can build reputation for arguing in a specific manner, or revealing their arguments in a particular order, for example, by presenting their cases to the same audience on a number of occasions. Then, what kind of reputation is in their interest to build? Our paper aims to offer some insight into this problem.

In economic theory, the main idea of reputation is that "if the player always plays in the same way, his opponents will come to expect him to play that way in the future and will adjust their play accordingly" (Fudenberg and Tirole, 1991, page 367). As such, we use this informal idea and capture reputation effects in our model by allowing the discussants to commit to strategies of their choice. The large literature on reputation supports this idea, and we conjecture that our model can be viewed as the reduced form of a traditional model of reputation effects arising from repeated interaction. However, the formal models of reputation are typically complicated, even in contexts much simpler than the present setting. It is thus beyond the scope of this paper to construct and analyze such a formal model.

Closely related to the commitment game is a model of costly information acquisition, which delivers another motivation for commitment. Imagine that agents must acquire signals before presenting them. The cost of acquiring one signal is negligible, but the cost of acquiring any additional signal is prohibitively large. There is a chance that an agent I searching for signal s_I or t_I will fail, i.e., signal $\neg s_I$ or $\neg t_I$, respectively, will be obtained. The audience observes which signal each agent tries to acquire as well as the evidence obtained. In Appendix B, we show that the results in this alternative model of costly information acquisition are similar to the results presented in the main text.[8]

5.3 Can the audience make debates more informative?

Although the objective of this paper is to derive some rules of thumb for discussants, it is also of interest to evaluate the results from the audience's perspective and ask whether the audience can do anything to create a more productive debate.[9]

A simple benchmark to consider is for the audience to request only one signal from each discussant sequentially, perhaps through a moderator who asks the discussants specific questions. The discussant has to reveal the requested signal if he has it, but reveals the other signal if that is the only signal he has. In effect, the audience chooses which signal (of the two) should be revealed if a discussant has both.

In the case of weak versus strong signals, the audience is clearly better off by requesting the strong signal from each discussant. This is different from the optimal commitment strategies for the discussants we derived in Propositions 1 and 2 under the Kahneman and Tversky preference. So the audience may benefit from intervention such as encouragement from the moderator of more aggressive argumentation.

In the case of independent versus correlated signals, recall from Proposition 5 that the leader

[8] Notice, however, that the two models are not equivalent. For example, an agent A who has both signals s_A and t_A, and contemplates which signal to reveal, has different information regarding agent B's signals than an agent A who is deciding which signal to search for.

[9] We would like to thank a referee for suggesting the questions addressed in this subsection and the next.

gives precedence to the correlated signal and from Proposition 4 that the follower responds to the correlated signal with the independent signal and to the independent signal with the correlated signal. This coincides with what the audience prefers – in particular, if the leader reveals a correlated signal, it is indeed better for the audience if the follower reveals whether he has the independent signal.

5.4 Timing of Commitment

We have so far assumed that the leader commits to a strategy first. This assumption of sequential commitment is reasonable in the model of costly information acquisition, discussed in Appendix B. However, under the interpretation of commitment as arising from reputation effects, there seems to be no reason for assuming any particular order in which agents commit to their strategies.

Fortunately, the equilibrium outcomes in the simultaneous commitment games coincide with the equilibrium outcomes of the sequential commitment game. More precisely, we can show the following results under the Kaheman and Tversky preference in the simultaneous commitment game: (1) In the weak versus strong case, it is a dominant strategy of the follower to respond with the strong signal to the strong signal revealed by the leader, and to respond with the weak signal to the weak signal revealed by the leader. The leader's best response to this strategy of the follower is to play the conciliatory strategy. (2) In the correlated versus independent case, the antagonistic strategy of the leader dominates the conciliatory strategy for ρ sufficiently close to 1. The follower's best response to the antagonistic strategy is to respond to a correlated signal with an independent signal, and to an independent signal with a correlated signal. These two results imply that the simultaneous commitment game has a unique equilibrium, and the outcome of this equilibrium coincides with the equilibrium outcome of the sequential commitment game. In this sense, our results are robust to different assumptions on the timing of commitment.

6 Conclusion

We provide a normative framework for the analysis of arguing in public debates. Each discussant's payoff depends on the audience's posterior belief in favor in his case. We focus on what we find to be most reasonable: the payoff is concave above the prior, convex below the prior, and concave at the prior. In this case, the model suggests that discussants should disregard arguments made by their opponents that do not seem relevant or convincing, and also respond with strong evidence to strong evidence. Moreover, they should begin a debate with presenting weaker rather than stronger evidence, although, as we show in Appendix A, these conclusions rely on the assumption that weak evidence is sufficiently week. The model also examines an explanation of the experimental findings of Glazer and Rubinstein (2001) based on the assumption that the correlation between the education levels of neighboring or similar cities is high.

We have studied only two applications of persuasion in this paper, but the model provides a framework for exploring other applications as well. One example is whether a discussant should build a reputation for speaking first or waiting until the opponent has made an argument. Another example is whether a discussant should build a reputation for presenting more original versus more standard evidence. That is, if the probability of obtaining signals s_A and s_B is relatively low in both states of nature; and the probability of obtaining signals t_A and t_B is higher in both states of nature, should the discussant reveal the more original evidence s_I or the more standard evidence t_I?

7 Appendix A: Strong versus Weak, but Informative, Evidence

In this appendix, we discuss how our analysis of weak versus strong evidence extends to the case in which the weak evidence is also informative about the state of nature. Formally, we assume $\rho = 0$ and $\varepsilon < \delta < 1/2$. By continuity, our results from the main text generalize to the case when δ is sufficiently close to 1/2. So, we are particularly interested in situations in which δ is not too close to 1/2. We restrict attention to the follower's problem; the leader's case turns out to be much less tractable.

Suppose first that the leader plays the conciliatory strategy and he presents the weak piece of evidence t_A. The follower has at most two pieces of evidence: a strong signal s_B and a weak signal t_B. The following table exhibits the posterior belief of the audience μ_b under each strategy of the follower, given the signals available to him:

	only t_B	only s_B	both s_B and t_B
str. t_B	$\mu_b = 1/2$	$\mu_b = \dfrac{(1-\varepsilon)\,\delta^2}{(1-\varepsilon)\,\delta^2 + \varepsilon\,(1-\delta)^2}$	$\mu_b = 1/2$
str. s_B	$\mu_b = \varepsilon$	$\mu_b = \dfrac{(1-\varepsilon)\,\delta}{(1-\varepsilon)\,\delta + \varepsilon\,(1-\delta)}$	$\mu_b = \dfrac{(1-\varepsilon)\,\delta}{(1-\varepsilon)\,\delta + \varepsilon\,(1-\delta)}$
	$\frac{1}{2}\delta\,(1-\delta)$	$\frac{1}{2}\left((1-\delta)^2\,\varepsilon + \delta^2\,(1-\varepsilon)\right)$	$\frac{1}{2}\delta\,(1-\delta)$

Table 10. The leader who plays the conciliatory strategy presented t_A.

Now suppose that the leader still plays the conciliatory strategy, but he presents the strong piece of evidence s_A. The following table exhibits μ_b under each strategy of the follower, given the signals available to him:

	only t_B	only s_B	both s_B and t_B
str. t_B	$\mu_b = \dfrac{(1-\delta)^2\,\varepsilon}{(1-\delta)^2\,\varepsilon + \delta^2\,(1-\varepsilon)}$	$\mu_b = 1/2$	$\mu_b = \dfrac{(1-\delta)^2\,\varepsilon}{(1-\delta)^2\,\varepsilon + \delta^2\,(1-\varepsilon)}$
str. s_B	$\mu_b = \dfrac{(1-\delta)^2\,\varepsilon^2}{(1-\delta)^2\,\varepsilon^2 + \delta^2\,(1-\varepsilon)^2}$	$\mu_b = 1-\delta$	$\mu_b = 1-\delta$
	$\dfrac{1}{2}\,(1-\varepsilon)^2\,\delta^2 + \frac{1}{2}\varepsilon^2\,(1-\delta)^2$	$\delta\,(1-\delta)\,\varepsilon\,(1-\varepsilon)$	$\dfrac{1}{2}\varepsilon\,(1-\varepsilon)\left(\delta^2 + (1-\delta)^2\right)$

Table 11. The leader who plays the conciliatory strategy presented s_A.

Proposition 6 ***(i)*** *Suppose the leader plays the conciliatory strategy. If* $\left(\frac{\delta}{1-\delta}\right)^2 \leq \left(\frac{\varepsilon}{1-\varepsilon}\right)$ *and the follower's utility is convex at* $1/2$*, convex on* $[0,1/2]$*, and concave on* $[1/2,1]$*, then the follower* *should respond to a weak signal with a strong signal. Also, if the follower's utility is strictly convex on* $[0, 1/2]$ *or strictly concave on* $[1/2, 1]$*, then the follower should respond to a weak signal with a strong signal.*

(ii) *Suppose the leader plays the conciliatory strategy and the follower's utility is concave at* $1/2$*, convex on* $[0, 1/2]$*, concave on* $[1/2, 1]$*. If* δ *is sufficiently close to* ε*, then the follower should respond to a strong signal with a weak signal.*

Proof: Part (i): As shown in Table 10, to compare the follower's utilities under strategies t_B and strategy s_B, we only need to compare two lotteries with the same mean. The lottery generated by strategy t_B has outcomes $1/2$ and $\frac{(1-\varepsilon)\delta^2}{(1-\varepsilon)\delta^2+\varepsilon(1-\delta)^2}$ (denoted by μ_1), which is lower than $1/2$ if $\left(\frac{\delta}{1-\delta}\right)^2 \leq \left(\frac{\varepsilon}{1-\varepsilon}\right)$; and the lottery generated by strategy s_B has outcomes ε (denoted by μ_2) and $\frac{(1-\varepsilon)\delta}{(1-\varepsilon)\delta+\varepsilon(1-\delta)}$ (denoted by μ_3), where

$$\mu_2 = \varepsilon < \mu_1 \leq 1/2 < \mu_3 < 1 - \mu_2$$

Since $1/2 < \mu_3 < 1 - \mu_2$, by the concavity of u_B at $1/2$ and its concavity on $[1/2, 1]$, the line passing through $(\mu_2, u_B(\mu_2))$ and $(\mu_3, u_B(\mu_3))$ is above the line passing through $(\mu_2, u_B(\mu_2))$ and $(1/2, u_B(1/2))$. By convexity on $[0, 1/2]$, the line passing through $(\mu_1, u_B(\mu_1))$ and $(1/2, u_B(1/2))$ is below than the line passing through $(\mu_2, u_B(\mu_2))$ and $(1/2, u_B(1/2))$. This yields that

$$E^{\text{str. } t_B}(u_B) \leq E^{\text{str. } s_B}(u_B).$$

Similarly, one obtains the strict inequality when u_B is strictly convex on $[0, 1/2]$ or strictly concave on $[1/2, 1]$.

Part (ii): As shown in Table 11, to compare the follower's utilities under strategies t_B and strategy s_B, we only need to compare two lotteries with the same mean. The lottery generated by strategy t_B has outcomes $1/2$ and $\mu_1 = \frac{(1-\delta)^2\varepsilon}{(1-\delta)^2\varepsilon+\delta^2(1-\varepsilon)}$, and the lottery generated by strategy s_B has outcomes $\mu_2 = \frac{(1-\delta)^2\varepsilon^2}{(1-\delta)^2\varepsilon^2+\delta^2(1-\varepsilon)^2}$ and $\mu_3 = 1 - \delta$.

When δ is sufficiently close to ε, we have

$$1 - \mu_3 = \delta \leq \mu_2 < 1/2 < \mu_1 < \mu_3.$$

Since u_B is convex on $[0, 1/2]$ and $1 - \mu_3 < \mu_2$, the line that goes through $(\mu_2, u_B(\mu_2))$ and $(1/2, u_B(1/2))$ is steeper than the line that goes through $(1 - \mu_3, u_B(\mu_3))$ and $(1/2, u_B(1/2))$. Since u_B is concave at $1/2$, it follows that the line that goes through $(\mu_2, u_B(\mu_2))$ and $(\mu_3, u_B(\mu_3))$ must be below the line that goes through $(1/2, u_B(1/2))$ and $(\mu_3, u_B(\mu_3))$. Also, since u_B is concave on

$[1/2, 1]$ and $\mu_1 < \mu_3$, the line that goes through $(1/2, u_B(1/2))$ and is $(\mu_1, u_B(\mu_1))$ is steeper than the line that goes through $(1/2, u_B(1/2))$ and is $(\mu_3, u_B(\mu_3))$. Hence, the line that goes through $(1/2, u_B(1/2))$ and is $(\mu_1, u_B(\mu_1))$ is above the line that goes through $(\mu_2, u_B(\mu_2))$ and $(\mu_3, u_B(\mu_3))$. This yields that

$$E^{\text{str. } t_B}(u_B) \geq E^{\text{str. } s_B}(u_B).$$

■

The advice for the follower in the case when δ is "small" is therefore different from that in the case when δ is close to 1/2. Indeed, It follows from Proposition 1 (i) that the follower should be indifferent between the two possible responses, when $\delta = 1/2$ and u_B is symmetric around 1/2, that is,

$$\frac{1}{2}u_B\left(\frac{1}{2}+x\right)+\frac{1}{2}u_B\left(\frac{1}{2}-x\right)=u_B\left(\frac{1}{2}\right), \text{ for every } x \in (0, 1/2],$$

independently of the shape of u_B on intervals $[0, 1/2]$ and $[1/2, 1]$. In contrast, Proposition 6 says that the follower should strictly prefer responding to a weak signal with a strong signal when δ is close to ε if his utility is strictly convex on $[0, 1/2]$ or strictly concave on $[1/2, 1]$, even when u_B is symmetric around 1/2.

Recall that strategy t_B generates a lottery with outcomes 1/2 and μ_1. An important difference comes from the fact that $\mu_1 \leq 1/2$ when $\left(\frac{\delta}{1-\delta}\right)^2 \leq \frac{\varepsilon}{1-\varepsilon}$ whereas $\mu_1 > 1/2$ for $\delta = 1/2$. Thus, curvatures of u_B in different regions matter for Propositions 1 (i) and Proposition 6 (i).

Note finally that the advice for the follower becomes ambiguous when the follower's utility is concave at 1/2. Informally speaking, there is a trade-off between the effects described in Propositions 1 (i) and 6 (i), and the advice for the follower depends on the convexity on $[0, 1/2]$ and the concavity on $[1/2, 1]$ compared to the concavity at 1/2.

If the leader plays the antagonistic strategy, then we get similar results as in Proposition 1 (ii) and (iii). That is, the results on the follower's optimal strategy generalize to the case in which the weak evidence is informative if the leader's strategy is antagonistic.

8 Appendix B: Costly Information Acquisition

In this appendix, we consider an alternative but closely related model of costly information acquisition. We assume that the cost of acquiring one signal is negligible, but the cost of acquiring another signal is prohibitively large. So, each agent acquires just one signal. Agent I searching for signal s_I

or t_I may fail, in which case signal $\neg s_I$ or $\neg t_I$, respectively, will be obtained. The signal obtained by the agents are publicly observed. Each agent decides what signal to acquire just before his turn to speak. In particular, the follower observes what the leader obtains before deciding what signal to acquire. The audience observes what signal each agent tries to acquire.

The model of costly information acquisition provides a robustness analysis for the results obtained in the main text. One feature of the model studied in the main text is that the audience must make inference from the strategies of the agents. That is, the audience's posterior depends not only on the arguments presented by the agents, but also on the audience's belief about their strategies. In contrast, the audience's posterior depends only on the presented arguments in the model of costly information acquisition.

We show that the results of the two models are consistent, but we also point out some differences.

8.1 Weak versus Strong Evidence

As in section 3.1, assume that $\rho = 0$ and $\varepsilon < \delta = 1/2$.

Suppose the leader's strategy is to acquire the weak signal t_A. Since $\delta = 1/2$, the audience's posterior is independent of what signal the leader obtains. If the follower decides to acquire the weak signal, then the posterior μ_b is $1/2$, no matter what signal the follower obtains. If the follower decides to acquire the strong signal s_B, then the posterior μ_b is ε if the follower obtains $\neg s_B$ and the posterior μ_b is $(1-\varepsilon)$ if the follower obtains s_B.

Suppose the leader's strategy is to acquire the strong signal s_A and he obtains s_A. If the follower decides to acquire the weak signal, then the posterior μ_b is ε, no matter what signal the the follower obtains. If the follower decides to acquire the strong signal, then the posterior μ_b is $\left(\frac{\varepsilon^2}{\varepsilon^2+(1-\varepsilon)^2}\right)$ if the follower obtains $\neg s_B$ and the posterior μ_b is $1/2$ if the follower obtains s_B.

Suppose the leader's strategy is to acquire the strong signal s_A and he obtains $\neg s_A$. If the follower decides to acquire the weak signal, then the posterior μ_b is $(1-\varepsilon)$, no matter what signal the the follower obtains. If the follower decides to acquire the strong signal, then the posterior μ_b is $1/2$ if the follower obtains $\neg s_B$ and the posterior μ_b is $\left(\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}\right)$ if the follower obtains s_B. To summarize, we have the following proposition.

Proposition 7 ***(i)*** *Suppose the leader plays the strategy of acquiring the weak signal. Independent of what the leader obtains, if the follower's utility is concave at* $1/2$*, he should acquire the weak signal; if the follower's utility is convex at* $1/2$*, he should acquire the strong signal.*

(ii) *Suppose the leader plays the strategy of acquiring the strong signal s_A and obtains s_A. If the follower's utility is convex on $[0,1/2]$, he should acquire the strong signal; if the follower's utility is concave on $[0,1/2]$, he should acquire the weak signal.*
(iii) *Suppose the leader plays the strategy of acquiring the strong signal and obtains $\neg s_A$. If the follower's utility is concave on $[1/2,1]$, he should acquire the weak signal; if the follower's utility is convex on $[1/2,1]$, he should acquire the strong signal.*

We now turn to the problem of the leader. As in section 3.2, we assume that the players' utility functions are convex on $[0,1/2]$, concave on $[1/2,1]$ and concave at $1/2$.

Suppose the leader plays the strategy of acquiring the weak signal. Then, as shown in Proposition 7 (i), the follower responds by acquiring the weak signal. So, the posterior μ_a is $1/2$ no matter what signals the players obtain.

Suppose the leader plays the strategy of acquiring the strong signal s_A. If the leader obtains s_A, then, as shown in Proposition 8 (ii), the follower responds by acquiring the strong signal and the posterior μ_a is $1/2$ if the follower obtains s_B and the posterior μ_a is $\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}$ if the follower obtains $\neg s_B$. If the leader obtains $\neg s_A$, then, as shown in Proposition 8 (iii), the follower responds by acquiring the weak signal and hence the posterior μ_a is ε no matter what signal the follower obtains. Since $\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2} > 1-\varepsilon$ and the leader's utility function is convex on $[0,1/2]$, concave on $[1/2,1]$ and concave at $1/2$, the line that goes through $(\varepsilon, u_A(\varepsilon))$ and $\left(\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}, u_A\left(\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}\right)\right)$ is below $(1/2, u_A(1/2))$. To summarize, we have the following result.

Proposition 8 *Suppose both players' utility functions are convex on $[0,1/2]$, concave on $[1/2,1]$, and concave at $1/2$. Then the leader should play the strategy of acquiring the weak signal t_A.*

8.2 Correlated versus Independent Evidence

As in section 4.1, assume that $\varepsilon = \delta < 1/2$ and $\rho > 0$. Suppose the leader's strategy is to acquire the independent signal t_A. Then, no matter what signal he obtains, the follower is indifferent between acquiring t_B and acquiring s_B.

Suppose the leader's strategy is to acquire the correlated signal s_A and he obtains s_A. If the follower decides to acquire the correlated signal s_B, then the posterior μ_b is $\left(\frac{\varepsilon^2+\rho\varepsilon(1-\varepsilon)}{\varepsilon^2+(1-\varepsilon)^2+2\rho\varepsilon(1-\varepsilon)}\right)$ if the follower obtains $\neg s_B$ and the posterior μ_b is $1/2$ if the follower obtains s_B. If the follower decides to acquire the independent signal t_B, then the posterior μ_b is $\left(\frac{\varepsilon^2}{\varepsilon^2+(1-\varepsilon)^2}\right)$ if the follower obtains $\neg t_B$

and the posterior μ_b is $1/2$ if the follower obtains t_B. Since $\frac{\varepsilon^2}{\varepsilon^2+(1-\varepsilon)^2} < \frac{\varepsilon^2+\rho\varepsilon(1-\varepsilon)}{\varepsilon^2+(1-\varepsilon)^2+2\rho\varepsilon(1-\varepsilon)} < 1/2$, there is more dispersion in the audience's posterior if the follower acquires the independent signal.

Suppose the leader's strategy is to acquire the correlated signal s_A and he obtains $\neg s_A$. If the follower decides to acquire the correlated signal s_B, then the posterior μ_b is $1/2$ if the follower obtains $\neg s_B$ and the posterior μ_b is $\left(\frac{(1-\varepsilon)^2+\rho\varepsilon(1-\varepsilon)}{\varepsilon^2+(1-\varepsilon)^2+2\rho\varepsilon(1-\varepsilon)}\right)$ if the follower obtains s_B. If the follower decides to acquire the independent signal t_B, then the posterior μ_b is $1/2$ if the follower obtains $\neg t_B$ and the posterior μ_b is $\left(\frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}\right)$ if the follower obtains t_B. Since $1/2 < \frac{(1-\varepsilon)^2+\rho\varepsilon(1-\varepsilon)}{\varepsilon^2+(1-\varepsilon)^2+2\rho\varepsilon(1-\varepsilon)} < \frac{(1-\varepsilon)^2}{\varepsilon^2+(1-\varepsilon)^2}$, there is more dispersion in the audience's posterior if the follower acquires the independent signal. To summarize, we have the following result.

Proposition 9 ***(i)*** *Suppose the leader plays the strategy of acquiring the correlated signal s_A and obtains s_A. If the follower's utility is convex on $[0, 1/2]$, then he should acquire the independent signal t_B. If the follower's utility is concave on $[0, 1/2]$, then he should acquire the correlated signal s_B.*

(ii) *Suppose the leader plays the strategy of acquiring the correlated signal s_A and obtains $\neg s_A$. If the follower's utility is convex on $[1/2, 1]$, then he should acquire the independent signal t_B. If the follower's utility is concave on $[1/2, 1]$, then he should acquire the correlated signal s_B.*

We now turn to the problem of the leader. As in section 4.2, assume that each agent's utility is convex on $[0, 1/2]$ and concave on $[1/2, 1]$. As shown in Proposition 10 (i), if the leader plays the strategy of acquiring the signal s_A and obtains s_A, then the follower responds by acquiring the independent signal t_B. If the leader plays the strategy of acquiring the signal t_A and obtains t_A, then the follower is indifferent between acquiring s_B and acquiring t_B. We can therefore without loss of generality assume that the follower responds by acquiring the signal t_B. Hence, the leader faces the same lotteries if he successfully obtains the signal that he decided to acquire, no matter what his strategy is. On the other hand, if the leader plays the strategy of acquiring the signal s_A and obtains $\neg s_A$, then, as shown in Proposition 10 (ii), the follower responds by acquiring the correlated signal s_B. The posterior μ_a is $1/2$ if the follower obtains $\neg s_B$ and the posterior μ_a is $\left(\frac{\varepsilon^2+\rho\varepsilon(1-\varepsilon)}{\varepsilon^2+(1-\varepsilon)^2+2\rho\varepsilon(1-\varepsilon)}\right)$ if the follower obtains s_B. If the leader plays the strategy of acquiring the signal t_A and obtains $\neg t_A$, the follower is indifferent between acquiring s_B and acquiring t_B. Again without loss of generality, we assume that the follower responds by acquiring the signal s_B. The posterior μ_a is $1/2$ if the follower obtains $\neg s_B$ and the posterior μ_a is $\left(\frac{\varepsilon^2}{\varepsilon^2+(1-\varepsilon)^2}\right)$ if the follower obtains s_B. Since $\frac{\varepsilon^2}{\varepsilon^2+(1-\varepsilon)^2} < \frac{\varepsilon^2+\rho\varepsilon(1-\varepsilon)}{\varepsilon^2+(1-\varepsilon)^2+2\rho\varepsilon(1-\varepsilon)} < 1/2$, the posterior μ_a has more dispersion on

$[0, 1/2]$ if the leader's strategy is to acquire the independent signal t_A.

Proposition 10 *Suppose both players' utility functions are convex on* $[0, 1/2]$ *and concave on* $[1/2, 1]$*. Then the leader should acquire the independent signal* t_A*.*

How can the Glazer-Rubinstein findings be reconciled with the model of costly information acquisition? Consider a scenario in which the leader happened to play the strategy of acquiring the correlated signal s_A, and obtains signal s_A. Then, Proposition 9 (i) says that the strategy s_B is preferred by an agent who is risk averse with respect to the audience posterior beliefs on interval $[0, 1/2]$, which is consistent with the conclusion from the main text.

Notice, however, that some conclusions of the costly acquisition model and the model from the main text are different. For example, they predict different strategies of the leader in the case when both agents have preferences which are concave above the prior, convex below the prior, exhibit some form of loss aversion, and the correlation coefficient is high.

9 Appendix C: Deriving Posterior in the Case of Correlated versus Independent Evidence

Tables $4', 5'$, $6'$ and $7'$ below exhibit the audience's posterior μ_a for different strategies of the leader. They contain the relevant rows from Tables 4, 5, 6 and 7.

	only s_B	only t_B	both s_B & t_B	neither s_B or t_B
str. t_B	$\mu_a = 1 - \varepsilon$	$\mu_a = 1/2$	$\mu_a = 1/2$	$\mu_a = \mu_4$
ex ante prob.	$\frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)$	$\frac{1}{2}(1+\rho)(1-\varepsilon)\varepsilon$	$\frac{1}{2}(1-\rho)\varepsilon(1-\varepsilon)$	$\frac{1}{2}(1-\varepsilon)^3$ $+\frac{1}{2}\varepsilon^3 + \frac{1}{2}\rho\varepsilon(1-\varepsilon)$

Table 4′. The leader who plays the antagonistic strategy has either s_A or both t_A and s_A.

So he presents s_A and the follower responds by strategy t_B.

	only s_B	only t_B	both s_B & t_B	neither s_B nor t_B
str. s_B	$\mu_a = \varepsilon$	$\mu_a = \mu_2$	$\mu_a = \varepsilon$	$\mu_a = \mu_3$
ex ante prob.	$(1-\rho)\varepsilon^2(1-\varepsilon)^2$	π	π''	π'

Table 5′. The leader who plays the conciliatory strategy has only s_A.

So he presents s_A and the follower responds with strategy s_B.

	only s_B	only t_B	both s_B & t_B	neither s_B nor t_B
str. s_B	$\mu_a = 1/2$	$\mu_a = 1 - \varepsilon$	$\mu_a = 1/2$	$\mu_a = \dfrac{(1-\varepsilon)^3}{(1-\varepsilon)^3 + \varepsilon^3}$
ex ante prob.	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}\varepsilon(1-\varepsilon)$	$\frac{1}{2}(1-\varepsilon)^3 + \frac{1}{2}\varepsilon^3$

Table 6′. The leader who plays the conciliatory strategy has either t_A or both t_A and s_A.

So he presents t_A and the follower is indifferent. We suppose he responds with s_B.

	only s_B	only t_B	both s_B & t_B	neither s_B nor t_B
str. s_B	$\mu_a = 1 - \mu_1$	$\mu_a = 1/2$	$\mu_a = 1 - \mu_1$	$\mu_a = \dfrac{(1-\varepsilon)^2}{\varepsilon^2 + (1-\varepsilon)^2}$
ex ante prob.	π	$(1-\rho)\varepsilon^2(1-\varepsilon)^2$	π'	π''

Table 7′. The leader who plays the antagonistic strategy has only t_A.

So he presents t_A and the follower responds by strategy s_B.

References

[1] Crawford, V., and J. Sobel (1982): "Strategic information transmission," *Econometrica*, **50**, 1431-1452.

[2] Dziuda, W. (2011): "Strategic Argumentation," *Journal of Economic Theory*, **146**, 1362-1397.

[3] Fudenberg, D., and J. Tirole (1991): *Game Theory*, MIT Press.

[4] Geanakoplos, J., D. Pearce and E. Stacchetti (1989), "Psychological Games and Sequential Rationality," *Games and Economic Behavior*, **1**, 60-79.

[5] Glazer, J., and A. Rubinstein (2001): "Debates and Decisions: On a Rationale of Argumentation Rules," *Games and Economic Behavior*, **36**, 158-173.

[6] Glazer, J., and A. Rubinstein (2004): "On the Optimal Rules of Persuasion," *Econometrica*, **72**, 1715-1736.

[7] Glazer, J., and A. Rubinstein (2006): "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach," *Theoretical Economics*, **1**, 395-410.

[8] Kahneman, D., and A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, **47**, 263-291.

[9] Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, **101**, 2590-2615.

[10] Milgrom, P. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, **12**, 350-391.

[11] Milgrom, P., and J. Roberts (1986): "Relying on the Information of Interested Parties," *Rand Journal of Economics*, **17**, 18-32.

[12] Olszewski, W. (2004): "Informal Communication," *Journal of Economic Theory*, **117**, 180-200.

[13] Sher, I. (2009): "Persuasion and Limited Communication," Unpublished Manuscript.

[14] Sher, I. (2011): "Credibility and Determinism in a Game of Persuasion," *Games and Economic Behavior*, **71**, 409-419.

[15] Shin, H-S. (1994): "The Burden of Proof in a Game of Persuasion," *Journal of Economic Theory*, **64**, 253-264.

[16] Thordal-Le Quement, M. (2010): "Persuasion and Rhetorical Moderateness," Unpublished Manuscript.