Influence Motivated Communication*

Konuray Mutluer†

CERGE-EI

This Version: 21 October 2025. Latest Version HERE

Abstract

I study strategic communication between a partially informed receiver and an informed expert who is motivated by influence: she values her advice being followed as an end in itself. The central result is that, despite lacking commitment power, the equilibrium behavior of the expert is identical to the optimal commitment solution of the corresponding Bayesian persuasion problem. Thus, a desire for influence provides an alternative microfoundation for this canonical framework when commitment is not feasible. Beyond this equivalence, I show that the receiver's level of informedness determines the direction of causality between advice and action: as the receiver becomes more informed, the expert ceases to be an "opinion leader" and becomes a "follower". The receiver gains the most from communication when she is neither too ignorant nor too well informed.

*I thank Matteo Escudé, Ole Jann, Frederic Koessler, David K. Levine, Filip Matějka, Laurent Mathevet, Andrea Mattozzi, Ran Spiegler, Yiman Sun, Jan Zapal and seminar participants at the SAET Meeting (Ischia), CERGE-EI, CUNEF, EUI, and UPV/EHU for useful comments.

†E-mail: konuray.mutluer@cerge-ei.cz

1 Introduction

Individuals often consult with an expert before making a "yes or no" decision in order to inform their choices. These experts may want a particular action to be taken regardless of what is best for the decision maker: an influencer may have a vested interest in selling the product, a political advisor may have personal stakes in passing a law.[1] In such settings, if advice is cheap talk (i.e., neither party has preferences over messages themselves), then communication is inconsequential for the action of the decision maker. Simply put, the expert is always tempted to switch to the advice most likely to induce her favored action. However, in many simple real-life decisions we observe such experts give consequential, albeit imperfect, advice. A prevalent approach is to suppose that the expert can commit to a messaging strategy before observing the state of the world (Kamenica and Gentzkow, 2011). This approach predicts reasonable expert behavior, but commitment can be difficult to justify in many environments.

This paper shows that an alternative to assuming commitment is to account for the expert's desire to be followed as an end in itself. Political advisors benefit from "having the government's ear", and influencers maintain their status by staying influential. I find that such an expert—even if she is free to deviate upon observing the state—advises in the same way as if she were driven purely by a bias towards a particular decision and could commit ex-ante. In other words, replacing commitment power by a desire to be followed preserves the behavioral predictions of standard persuasion models. Moreover, this new framework provides insights about the nature of causality between advice and action.

To this end, I model one-shot strategic communication between an expert and a consumer about the quality of a product, which is either high or low.[2] The expert knows the state (i.e., quality) and sends either a positive or a negative review. The consumer observes the review and, importantly, her own private "opinion" about the quality (i.e., a noisy private signal

[1] Such experts are referred to as ones with "transparent motives" in the literature (e.g., Lipnowski and Ravid, 2020). The environment itself is referred to as "no common interest" between the expert and the decision maker.

[2] Throughout the paper I stay with the product/consumer case for expositional clarity. The analysis applies to any setting with the above described features.

about the state). Upon these observations, the consumer makes the decision of whether to buy or reject. The consumer simply prefers to buy if the quality is high, and reject otherwise. The preferences of the expert have two additive elements. First is a simple bias: she enjoys some utility if the consumer buys. Second, she wants the consumer's decision to match the advice. That is, she enjoys some utility if the outcome is either "buy" with "positive review", or "reject" with "negative review". I refer to the second component as "influence motivation". This expert has transparent motives, as she does not care about quality; her messages are not cheap talk, as her utility depends on the message itself.

This game yields a unique informative equilibrium if the influence motivation is sufficiently strong relative to the bias.[3] In this equilibrium, the expert gives a positive review for certain if the quality is high, and randomizes otherwise. The consumer follows the review for certain except in one case: when the review is positive but her private opinion is negative. In this case, she randomizes between buying and rejecting.

The main result shows that the expert's equilibrium behavior is identical to the solution of a standard Bayesian persuasion problem in the same environment. Specifically, I compare the unique informative equilibrium of my model to the optimal review strategy of an expert who can ex-ante commit. In this commitment benchmark, the expert is purely driven by a bias toward inducing the consumer to buy. Therefore, it effectively replaces the influence motivation by commitment power. The comparison yields the following: both kinds of experts give one review for certain if the quality is high and mix with identical probabilities otherwise.[4] Hence, we may observe such advice not because experts can commit, but because they care about being followed.

The intuition behind this behavioral equivalence is that both experts face the same trade-off: they aim to send the review that induces the consumer to buy as often as possible, but not so often that it is optimally ignored. With commitment, Bayes-Plausibility enforces this constraint. With influence motivation, it is enforced by the direct utility punishment from being ignored. This mechanism is not an artifact of the simple "two-action-two state" model,

[3]Otherwise (i.e., if the influence motivation is very weak) communication is futile, and all equilibria are in pooling strategies by the expert.

[4]The commitment problem I consider does not include reviews that depend on the private signal realization of the consumer. The domain of a commitment strategy is only the states of the world.

and I show this by deriving conditions under which expert behavior remains equivalent under an arbitrary number of states and actions.

The behavior of the consumer is different in the two cases: in the commitment model the consumer always does as the review suggests, whereas in the influence motivated model she may reject despite a positive review.[5] However, as the bias of the influence motivated expert goes to zero (i.e., in the limit where she is "purely influence motivated"), consumer's probability of rejecting after a positive review disappears, and her behavior converges to the commitment case.

The comparison of the influence motivated and commitment models further suggests a self-reinforcing mechanism through which message labels acquire their meaning as state representations: under commitment, either review (positive or negative) can serve as a signal for either quality level. Under influence motivation, by contrast, the positive (negative) review must indicate high (low) quality. In the latter case each review label is fixed to a particular state because (*i*) the expert desires a particular mapping between review and consumer action, and (*ii*) this mapping is confirmed by the optimal response of the receiver only when each review stems from the "right" state.

Next, I demonstrate that when the expert is influence motivated, the direction of causality between advice and action hinges on the consumer's level of informedness. If the consumer's private signal is inaccurate, then the advice by the expert faces little resistance and she can easily persuade the consumer to buy. This way, the expert is effectively "leading" the interaction. As the consumer becomes more informed, however, two things change. First, persuasion becomes more "costly", as it requires outweighing the consumer's signal. Second, the consumer's signal becomes more correlated with the true state, and therefore its realization is more predictable. This makes it easier for the expert to align her advice with the consumer's intended action. Thus, when the consumer is well informed, the expert stops "telling her what to do" and starts echoing what the consumer would "do anyway". The causality of communication reverses itself: The expert ceases to be a leader and becomes a follower.

[5] Imposing the "expert-preferred" action by the consumer is a common requirement for existence of an optimal commitment strategy in Bayesian Persuasion literature.

The comparative statics with respect to consumer informedness yield notable insights on her behavior and welfare. First, the consumer's gain from communication relative to autarky (i.e., relying solely on her own signal) is non-monotonic in the accuracy of her private signal. It is maximized at a unique interior accuracy level, and is zero in the two extremes—where she has no information and perfect information. Therefore, if the consumer wishes to maximize the benefits from advice, she should be neither too ignorant nor too well informed when engaging with the expert. The interior optimum of private information has important implications for the optimal number and type of experts one should consult, which I discuss in detail.

Second, if the signal of the consumer is made public, she loses all gains from communication. In this case the expert merely echoes the (common knowledge) opinion of the consumer and her advice contains no information. The consumer's utility is the same as under autarky. Hence, when the expert and consumer have no common interest, the entire value of communication stems from the privacy of the consumer's information (or, more generally, her type). This observation provides a strong argument for the privacy of personal history and preference data.

Relation to the Literature

I depart from the common approach of cheap talk in strategic communication (Crawford and Sobel, 1982) by considering a sender who cares about the messages themselves. The focus is the value of communication when she has no direct preferences over the state of the world (i.e., transparent motives). Previous studies show that in more complex settings, a sender with transparent motives can be informative through cheap talk messages. For instance, this is shown to be the case when the state is multi-dimensional (Chakraborty and Harbaugh, 2010) or when there are several actions and the status-quo is not preferable for the sender (Lipnowski and Ravid, 2020).[6] In Dziuda and Salas (2018), informative cheap talk is possible when there is a chance that false messages are detected by the receiver. I show that by accounting for a natural concern (i.e., influence motivation), we can achieve meaningful

[6] This latter work characterizes the set of payoffs that are obtainable via cheap talk when the sender has transparent motives.

communication by such experts in even the simplest of environments. Furthermore, while these studies rely on the sender's indifference under all states, I show that state dependent (strict) preferences can arise as an equilibrium phenomenon.

I compare my equilibrium behavior to the predictions of another prominent approach where the sender can commit ex-ante to a messaging strategy (Kamenica and Gentzkow, 2011). In particular, replacing my influence motivation by the ability to commit corresponds to Bayesian Persuasion with a privately informed receiver (e.g., Kolotilin et. al., 2017; Kolotilin, 2018). I show that the predicted behavior is remarkably similar (identical for the sender) in the two cases. This suggests that such patterns of communication can still arise in settings where commitment is not possible.

Due to the important role played by the receiver's "opinion", this paper is related to work on strategic communication with an imperfectly informed receiver (Chen, 2012; Ishida and Shimizu, 2016; Lai, 2014; Moreno de Barreda, 2024). These studies consider cheap talk and rely on some alignment between the state dependent preferences of the sender and receiver (as opposed to the transparent motives of my sender).[7] My observations on the causality of communication and receiver's non-monotonic gain, on the other hand, stem from the tension between sender's preference to be followed and the potentially contradictory private opinion of the receiver. Therefore, the analysis provides new insights on the effects of consumer informedness on the nature of advice-giving.

In addition to intrinsic value or direct monetary incentives, one of the possible sources of influence motivation is to maintain a reputation as a consequential expert. In this sense, this paper is related to studies of information transmission with reputation-building. These include a reputation for "truth-telling" (Olszewski, 2004; Mathevet et. al., 2022) or a reputation for "expertise" (Ottaviani and Sørensen, 2006). These types of reputation, however, are in the eyes of the receiver. Thus, we can think of them as "downstream" forms of reputation. The influence reputation I consider would provide benefits when formed from the perspective of third-parties (such as lobbyists going to a political advisor or firms giving endorsement deals to influencers). Therefore, one way to think of influence motivation is as an "upstream"

[7] In particular, Moreno de Barreda (2024) shows that if this is the case, and if the sender is risk averse, then a more informed receiver can mean less informative communication and even lower ex-ante welfare.

reputation concern. Influence motivation is also related to a concern considered by Gallice and Grillo (2025) in a different setting, where a politician can endorse a norm change (a parameter in a coordination game played among the population), but suffers blowback if it does not catch on.[8]

Finally, while I use strategies for the formal derivation of the results, I use a belief-based approach in their exposition and discussion. This approach is related to the one used in recent work by Boleslavsky and Shadmehr (2024), and Koessler et. al. (2024) for general signaling games with an *uninformed* receiver. These two studies characterize equilibrium sender payoffs with and without commitment, respectively. My discussion complements theirs by incorporating a privately informed receiver into the belief based analysis of signaling.

The rest of the paper is structured as follows. Section 2 describes the model. Section 3 states and discusses the equilibria. Section 4 compares the equilibrium behavior to communication under commitment and pure bias. Section 5 discusses the effects of receiver informedness and the causal link between advice and action. Section 6 extends the model to allow an arbitrary number of states and actions. In this generalized setting, I derive a sufficient condition for the behavioral equivalence between commitment and influence motivation. Section 7 concludes.

2 Model Setup

An expert E and a decision maker D play a signaling game over a product of uncertain quality, determined by state $\omega \in \{0, 1\}$ (bad or good). First, the expert (perfectly) observes the realization of ω and sends the decision maker a message $m \in \{0, 1\}$, where $m = 0$ and $m = 1$ corresponds to a "negative" and "positive" review, respectively. Second, the decision maker observes message m as well as the realization of a noisy private signal $s \in \{0, 1\}$, which represents her own "opinion".[9] Finally, the decision maker chooses an action $a \in \{0, 1\}$

[8]Their different setting, as well as the differences in sender motivation leads to different patterns of behavior: In their model, the popularity concern is "one sided", in that being followed only matters if sender plays a certain action. Furthermore, the sender does not have a "bias" over the actions of the population.

[9]There are alternative ways of interpreting the private signal s of D, which are discussed below along with the corresponding comparative statics.

(reject or buy).

The players share a common prior $Pr(\omega = 1) = Pr(\omega = 0) = 0.5$.[10] The accuracy of D's private signal is $p := Pr(s = 1|\omega = 1) = Pr(s = 0|\omega = 0)$ with $p \in (0.5, 1)$. While the realization of s is only observed by the decision maker, its conditional distribution is common knowledge.

The decision maker's payoff depends on her action a and state ω. In particular, she receives utility 1 if she buys a good product, -1 if she buys a bad product, and 0 if she rejects. We can summarize her utility function as follows.[11]

$$u_D(\omega, a) = \mathbb{1}\{(a, \omega) = (1, 1)\} - \mathbb{1}\{(a, \omega) = (1, 0)\}$$

The expert's payoff is determined by her message m and D's action a. In particular, she receives utility from two sources: First, she receives utility $\beta > 0$ if $a = 1$. This is her pure motivation to sell the product. Henceforth I use the term "bias" to refer to this component of E's preferences. Second, E receives an additional utility of 1 if $a = m$. Therefore E values that the decision of D is in line with her advice m. Henceforth, I refer to this component as the "influence motivation". We can summarize E's utility as follows.

$$u_E(m, a) = \beta a + \mathbb{1}\{a = m\}$$

It is useful to emphasize that E's preferences are independent of ω. Therefore, this is a setting involving transparent motives (also referred to as "no common interest" in the literature). The solution concept is Perfect Bayesian Equilibrium.

3 Equilibrium

In this section, I first state and interpret the equilibria. Second, I outline the argument behind the informative equilibrium.

[10] The assumption of equal priors is made purely for expositional clarity. All observations carry through with sufficiently moderate priors.

[11] Similar to the uniform prior assumption, the normalization of D's payoffs (0 for $a = 0$, symmetric values for $a = 1$) is not required for the analysis. All results extend to any specification where D strictly prefers to match the state, i.e., where her unique optimal action under state ω is $a = \omega$. This is illustrated in the extension of Section 6, where D minimizes a quadratic loss function.

A mixed strategy of E is given by function $\sigma_E : \{0,1\} \mapsto \Delta\{0,1\}$, which specifies a distribution over possible messages for each realization of ω. For conciseness, denote by σ^ω the probability that E sends $m = 1$ when the state is ω. Next, note that D has to make a decision under four information sets: for each message sent by E, and for each realization of her own private signal. Therefore, her mixed strategy is given by a function $\sigma_D : \{0,1\}^2 \mapsto \Delta\{0,1\}$. Denote by $\sigma^{m,s}$ the probability that D plays $a = 1$ upon receiving message m and private signal s. In what follows, I state the mixed strategies for E as $\sigma_E := (\sigma^1, \sigma^0)$ and for D as $\sigma_D = (\sigma^{1,1}, \sigma^{1,0}, \sigma^{0,1}, \sigma^{0,0})$.

The focus of the analysis is equilibria in strategies such that $\sigma_E \notin \{(1,1),(0,0)\}$. That is, I only look at strategies where E sends each message with positive probability under some state. Clearly, this is a necessary condition for informative communication, in the sense that observing m leads to an on-path change in D's belief about the state. This restriction still allows for babbling equilibria, and even equilibria where E plays a mixed pooling strategy.

3.1 Statement and Interpretation

The following theorem states the equilibria of the game. Complete proofs can be found in the Appendix.

Theorem 1. *If $\beta < \frac{p}{1-p}$, then there is a unique equilibrium that satisfies $\sigma_E^* \notin \{(1,1),(0,0)\}$. The strategies are as follows.*

$$\sigma_E^* = \left(1, \frac{1-p}{p}\right) \quad \wedge \quad \sigma_D^* = \left(1, \frac{p(1+\beta)-\beta}{p(1+\beta)}, 0, 0\right)$$

Otherwise, $\sigma_E^* \in \{(1,1),(0,0)\}$ in any equilibrium.

Recall that the influence motivation component of E's utility is normalized to 1, and therefore a smaller bias β means stronger (relative) influence motivation. We can summarize the result as follows. If E is sufficiently influence motivated, then there exists a unique equilibrium in non-pooling strategies. In this equilibrium, E sends $m = 1$ for certain if $\omega = 1$ and with probability $\frac{1-p}{p}$ if $\omega = 0$. Therefore, E "over-recommends" the product, in the sense that she always gives it a positive review if it is of high quality and randomizes otherwise. Decision maker D plays $a = 0$ for certain following $m = 0$ regardless of the realization of s, plays $a = 1$ for certain if $m = s = 1$, and randomizes if $m = 1$ but $s = 0$. That is, if E gives the product a bad review, D follows it regardless of her own opinion. The response to a positive review depends on D's own opinion: It is followed for certain if D's opinion is also positive. If the positive review conflicts D's negative opinion, then D randomizes.

Regardless of E's preferences, this game always yields equilibria with $\sigma_E \in \{(1,1),(0,0)\}$. For instance, $\sigma_E = (0,0)$ is always supported in equilibrium by the off-path belief that the product is low quality for certain, which means a deviation results in the lowest feasible payoff in the game for E. If $\beta > \frac{p}{1-p}$, then all equilibria of the game are in pure pooling strategies.

It is important to underline that although the utility of E is independent of the true state (i.e., she has transparent motives), she can be informative in equilibrium. Unless the private signal of D and the influence motivation of E are both present, transparent motives imply that E's optimal message carries over across different states for any strategy of D. Then, equilibrium relies on E's indifference under all states.[12] For our setting, this would mean that any equilibrium is babbling.

With influence motivation and privately informed D, however, state dependent optimal messages arise as an equilibrium phenomenon: First, the true state is correlated with the private signal of D. This means for given strategy of D, different states provide E with different forecasts about the expected reaction to each message. Second, equilibrium informativeness is not possible even when D is privately informed but E is not influence motivated (i.e., she only gains utility from $a = 1$). In that case, E would always have incentive to send whichever message makes D most likely to buy. For any message m, however, D is (weakly) more likely to buy under $s = 1$ than under $s = 0$. This means the message that leads to $a = 1$ with the highest probability is independent of state. The combination of these two considerations, on the other hand, makes the incentives of E state dependent. Indeed, in the informative equilibrium stated in Theorem 1, E strictly prefers $m = 1$ under $\omega = 1$ and is indifferent under $\omega = 0$.

[12] This requirement in the case with cheap talk messages and an uninformed receiver is discussed by Lipnowski and Ravid (2020).

3.2 Informative Equilibrium Structure

While the formal proof of Theorem 1 uses strategy profiles, it is useful to directly consider interim beliefs to illustrate the structure of the informative equilibrium. Here, "interim" refers to D's beliefs about ω after updating based on message m, but before the realization of private signal s.[13] Denote by μ^m the interim belief that $\omega = 1$ upon observing message m. Assume that D plays her best response to all realizations of (m, s), and denote by $v(\mu^m|\omega)$ the expected utility of E from inducing μ^m via message m under state ω. The expectation is taken over two random variables: First, for given (m, s), E is unsure about the realized action from a potentially mixed strategy by D. Second, E is unsure about the realization of s itself. This latter factor makes v state dependent, since ω contains information about s.

Figure 1 demonstrates the expected utility of E as a function of interim beliefs. Below the x-Axis is a description of the optimal receiver response to the corresponding interim belief. Note that if m is a sufficiently strong indicator of state ω (i.e., $\mu^m \notin [1-p, p]$), then D ignores her own signal s and take action $a = \omega$. If m is a weak indicator of states (i.e., $\mu^m \in (1-p, p)$), then D ignores m and follows her own signal s.

In equilibrium, any message that is sent with positive probability must be weakly preferred. That is, $\mu^m > 0$ and $\mu^m < 1$ has to imply $v(\mu^m|1) \geq v(\mu^{-m}|1)$ and $v(\mu^m|0) \geq v(\mu^{-m}|0)$ respectively.[14] Using these conditions, we can construct the unique "non-pooling" equilibrium in four steps. First, we must have $\mu^1 \in [1-p, p]$: Otherwise, sending $m = 1$ gives E either the strictly highest or lowest feasible payoff ($1+\beta$ and 0 respectively), yielding incentive to deviate in either case. Second, we must have $\mu^0 \in \{0, 1\}$: From Step 1, we know that μ^1 is interior, which means $m = 1$ is sent with positive probability under both states. If μ^0 is also interior, this would require that E is indifferent under both states. However, this is not possible because in the belief region $[1-p, p]$ (where μ^1 has to lie) the payoff of E is

[13] In a recent working paper, Koessler et. al., (2024) use a related approach in order to characterize the equilibrium sender payoffs for a general class of signaling games with an *uninformed receiver*. The beliefs of an uninformed receiver, however, are fully formed after observing only the message by the sender (here m). Therefore, they construct their approach using posterior beliefs.

[14] This is a necessary incentive compatibility condition. For sufficiency, we additionally need Bayes-Plausibility. That is, $\tau\mu^1 + (1-\tau)\mu^0 = 0.5$, where τ denotes the ex-ante probability of sending $m = 1$ according to strategy σ_E.

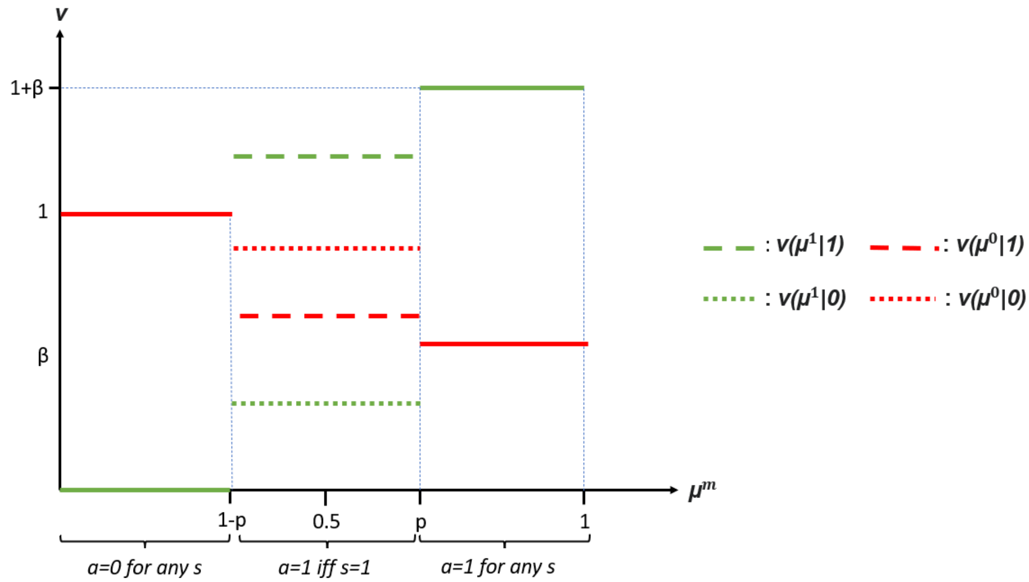

Figure 1: Expected payoffs from sending message m under state ω, as a function of induced interim belief μ^m. The green and red lines correspond to $m = 1$ and $m = 0$ respectively. The dashed and dotted lines correspond to $\omega = 1$ and $\omega = 0$ respectively. The payoff is state independent for $\mu^m < 1 - p$ and $\mu^m > p$, represented by the solid lines. Below the x-Axis, the optimal response for D to belief μ^m is indicated.

state dependent. Therefore, indifference in one state implies a strict preference in the other.

Third, we must have $\mu^0 \neq 1$: To have $\mu^0 = 1$ would require a weak preference for $m = 1$ under $\omega = 0$ (since $m = 1$ would be the only message sent in that state). However, since $\mu^1 \in [1 - p, p]$, the payoff from playing $m = 1$ is strictly higher under $\omega = 1$ than under $\omega = 0$ (as D is more likely to receive $s = 1$ and buy under $\omega = 1$). Then, a weak preference for $m = 1$ under $\omega = 0$ implies a strict preference for $m = 1$ under $\omega = 1$. This contradicts the optimality of sending $m = 0$ under $\omega = 1$, which is a necessary condition for $\mu^1 = 1$.

Hence, the remaining candidates for equilibrium beliefs must satisfy $\mu^0 = 0$ and $\mu^1 \in [1 - p, p]$. Due to Bayes-Plausibility, we can restrict our attention to $\mu^1 \in (0.5, p]$. That is, $m = 0$ perfectly reveals $\omega = 0$, and $m = 1$ signals $\omega = 1$ weakly enough that rejecting is still a best response for D under $s = 0$. For such a messaging strategy to be incentive compatible, E must weakly prefer $m = 0$ under $\omega = 0$, and $m = 1$ under $\omega = 1$. This is only possible when $\mu^1 = p$. In this case, D is indifferent between buying and rejecting when she observes $m = 1$ and $s = 0$. Then she can randomize in a way that gives E utility 1 and makes her indifferent between the two messages under $\omega = 0$. With this response, E strictly prefers

$m = 1$ under $\omega = 1$, so incentive compatibility is satisfied. Figure 2 demonstrates the interim beliefs and the resulting expected payoffs for E in the unique informative equilibrium.

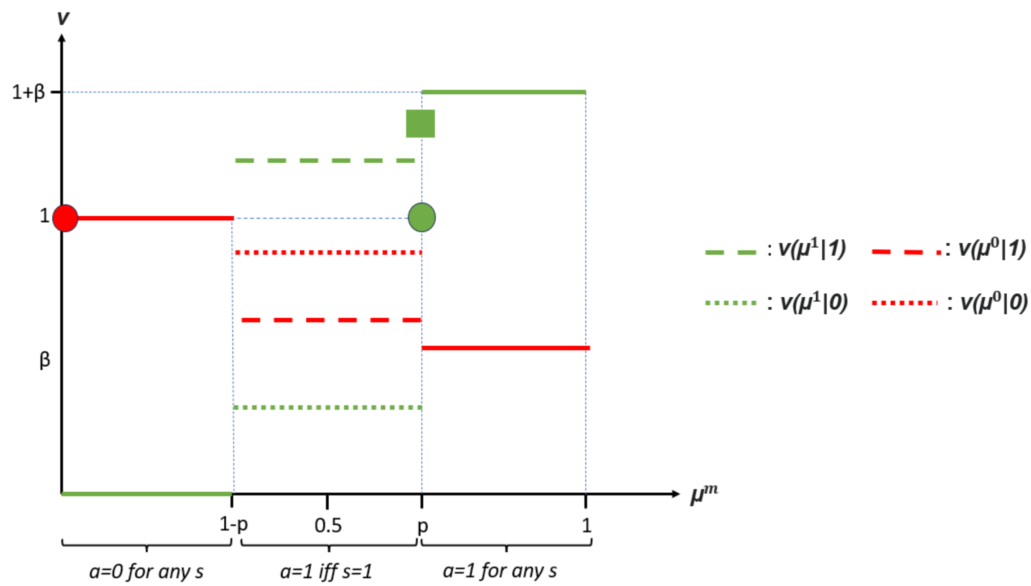

Figure 2: The unique informative equilibrium. The circles and the square denote E's payoff under $\omega = 0$ and $\omega = 1$ respectively. A negative and a positive review induce belief $\mu^0 = 0$ and $\mu^1 = p$ respectively.

To summarize, the only non-pooling strategy of E that is incentive compatible under optimal response is one that makes D indifferent when the positive review is met by a negative private opinion. This is achieved by giving a positive review for certain if the quality is high, and mixing in a way such that $m = 1$ "cancels out" $s = 0$ if quality is low.

4 Commitment with Pure Bias

Next, I compare my influence-motivated equilibrium to behavior resulting from a common approach in the analysis of information transmission. Consider two modifications to the model discussed so far: First, suppose E commits to a binary messaging strategy before observing state ω (à la Kamenica & Gentzkow, 2011) and cannot deviate afterwards.[15]

Second, suppose E only cares about selling and does not have the influence motivation. That is, $u_E(m, a) = \beta a$. With this specification the messages are cheap talk, as neither player has preferences over them.[16]

All else is the same as described in Section 2. Thus, we are effectively replacing E's influence motivation by the ability to ex-ante commit. This yields a standard model of Bayesian Persuasion with the addition of a partially informed receiver.[17]

The comparison is of particular interest because the commitment approach predicts intuitive and reasonable sender behavior, even in environments where the commitment assumption itself does not seem a good fit: Biased experts (e.g., political consultants or product reviewers) can often privately change their minds about their recommendation after they observe the relevant state of the world.

4.1 Optimal Commitment Outcome

The following result describes the behavior of E under the optimal commitment equilibrium.

Theorem 2. *There exists $\bar{p} \in (0.5, 1)$ such that for $p < \bar{p}$, the message structure in the optimal commitment equilibrium is as follows: E sends one message $m' \in \{0, 1\}$ for certain under $\omega = 1$, and with probability $\frac{1-p}{p}$ under $\omega = 0$. She sends the other message $m'' \in \{0, 1\} \setminus \{m'\}$ with the complementary probabilities.*

[15] It is without loss of generality to reduce the set of possible messages in the commitment problem to $m \in M := \{0, 1\}$. The "revelation" result (Proposition 1) of Kamenica & Gentzkow (2011) allows restricting attention to the set of "straightforward messages" $M \subseteq A$, where A is the set of possible actions. While this result is established for an uninformed D, it extends to the private information game when D's action set is binary. In this case, the expected value of D's best response action—taken over realizations of her private signal s—is a sufficient statistic to elicit the value of E from inducing an interim belief. We can therefore use the same approach as in the uninformed case, taking the expected action wherever D's action is sensitive to her private signal.

[16] Here, the term "cheap talk messages" describes only the property that message labels are payoff irrelevant. This is not to be confused with the complete cheap talk setting that is prevalent in the literature, which does not include the commitment ability.

[17] For such a model, see Kolotilin (2018).

The main takeaway from Theorem 2 is that fixing $m' = 1$ and $m'' = 0$ yields precisely the unique non-pooling equilibrium strategy σ_E^* of the influence motivated game.

More specifically, when the private signal of D is not too strong (i.e., low p), E "over-recommends" as in the influence-motivated equilibrium: In the high quality state, she sends one of the messages for certain. In the low quality state, she mixes. The mixture is the same as in the original model where E cannot commit but is influence motivated. In terms of information revealed, one message perfectly reveals low quality and the other imperfectly signals high quality (in a way that makes D indifferent after $s = 0$).

The only D behavior that supports this equilibrium is one that chooses E's favorite action when indifferent. Consequently, D buys for certain after the message that signals the high state and rejects for certain otherwise. In other words, she completely ignores her own private information.

To construct the equilibrium, we once again consider the expected utility of E as a function of D's interim beliefs, assuming optimal response. Denote this by $v(\mu)$. The commitment problem is represented in Figure 3.

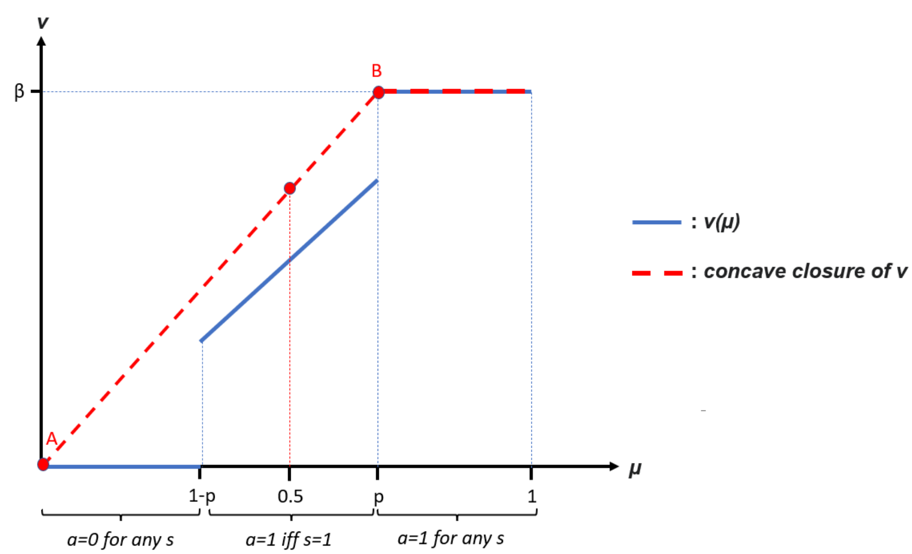

Figure 3: The commitment problem of a "purely biased" E given $p < \bar{p}$. The blue curve is the expected value $v(\mu)$ of E from inducing interim belief μ. For $\mu \in (1-p, p)$, the expectation is taken over s, and $\mu \notin (1-p, p)$, it is deterministic. The red (dotted) curve is the concave closure of $v(\mu)$. Points A and B indicate the interim beliefs and the value from the two messages under optimal commitment strategy. Note that the optimal belief pair $(0, p)$ is identical to the influence motivated equilibrium σ_E^*.

Function $v(\mu)$ differs from the case with influence motivation and no commitment in two ways: First, without influence motivation the messages are payoff irrelevant. In this sense, they are "cheap talk". Therefore, E's expected utility from inducing a belief is independent of which message she uses to induce it. This can be seen in Figure 3 from the value consisting of a single "m-independent" curve. Second, E makes her commitment decision before observing the state. Consequently, she cannot use her observation to make predictions about the private signal and the response of D. She has to use the common prior to predict these. This makes her expected utility state independent. The result is a simpler problem where the expected utility can be described by a single curve.

The optimal commitment strategy is determined by a pair of interim beliefs μ that satisfy the following condition: The Bayes-plausible convex combination of the utilities resulting from the two beliefs must lie on the concave closure of the value function $v(\mu)$. The unique pair of beliefs that satisfies this is $(0, p)$. This pair of interim beliefs, in turn, is obtained through a unique message structure. This message structure is one with a "dump message" that is sent only under $\omega = 0$, and another with probabilities such that it exactly induces interim belief p.

4.2 Commitment versus Influence Motivation: Expert

It is useful to emphasize the similar behavior of E in the case where she is influence motivated and the case when she can commit. As can be seen from Theorems 1 and 2, the equilibrium strategy of E is essentially identical in the two cases when D is not too well informed: One message reveals low quality, and the other makes D indifferent when there is a "conflict" between her own opinion and E's review. The only difference is that in the commitment case either message can play either role. Under influence motivation, however, it has to be the positive review that indicates the high state. This suggests a self-reinforcing way in which message labels obtain their meanings: E cares about the mapping between messages and actions, so she has to send the "right" message in the "right" state. This mapping is confirmed by the optimal response from D.

The intuition behind the equivalence is simple. In both models, E chooses messages in the same manner: She wants to send the message that indicates the state ($\omega = 1$) that induces her favorite action ($a = 1$) as often as she can, subject to the constraint that her message is not rationally ignored by D. With commitment, this constraint is enforced by Bayes Plausibility. In the influence motivated model it is enforced by the incentive compatibility requirement in each state, since E suffers a direct utility loss from "being ignored".

An important takeaway from the comparison of the two models is the following: If we replace the ability to commit by influence motivation, we can predict the same over-recommendation behavior by biased experts. Thus, this reasonable pattern of communication may stem from the fact that experts care about being followed, instead of their ability to commit.

4.3 Commitment versus Influence Motivation: Decision Maker

Turning to the equilibrium behavior of D, we see that she follows E's message for certain in both models except at one information set: When the review is positive ($m = 1$) but D's own opinion is negative ($s = 0$). With a purely biased expert who can commit, she buys for certain after this history as well, which means she always follows m regardless of the realization of s. This restriction, and more generally imposing that D does what E wants when indifferent, is a common requirement in Bayesian Persuasion for the existence of an optimal messaging strategy.

In the influence motivated model, on the other hand, D randomizes between following the positive review from E and following her own negative opinion. She does so in a way that makes E indifferent between "lying" and "telling the truth" when the product is low quality. Behavior of D therefore depends on the degree of bias of E. Specifically, the higher the weight of influence motivation in E's preferences (i.e., lower β), the more likely that D follows her advice (i.e., higher $\sigma^{1,0}$). Moreover, we can make the following observation.

Remark 1. *In the unique informative equilibrium of the influence motivated model, we have* $\lim_{\beta \to 0} \sigma_D^* = (1, 1, 0, 0)$. *That is, as E becomes purely influence motivated, the equilibrium strategy of D converges to her behavior under optimal commitment with pure bias: She follows m for certain and ignores s.*

Hence, the D behavior that is required to sustain the commitment equilibrium can be generated without commitment, as a limit case where E's only concern is that D follows her advice (regardless of what sort of advice it is).

5 Decision Maker's Informedness and Welfare

Returning to the influence motivated model, I now investigate the effects of D's private signal accuracy p. First, I look at how it affects behavior. Then I study D's welfare and her gains from communication. Finally, I briefly discuss alternative interpretations of p and the corresponding takeaways.

5.1 Signal Accuracy and Equilibrium Behavior

Figure 4 demonstrates the strategies (σ_E^*, σ_D^*) in the unique informative equilibrium as a function of p.

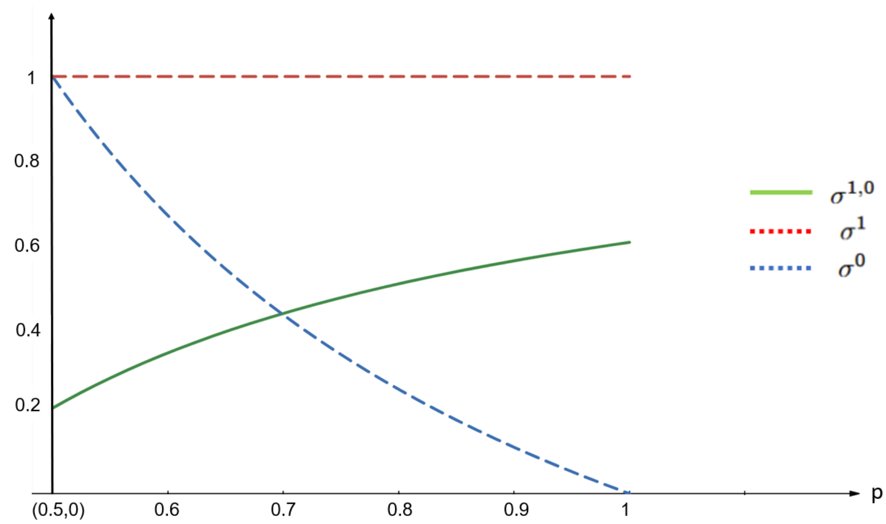

Figure 4: Equilibrium strategies (of the influence motivated model) as a function of p, given $\beta = 2/3$. The red and blue dashed lines represent the prob. of $m = 1$ under $\omega = 1$ and $\omega = 0$ respectively. The solid green line is the probability of $a = 1$ under $m = 1$ and $s = 0$. In all other information sets, D plays $a = m$ for certain.

The behavior of E is rather intuitive. She always gives a positive review if the quality is high. It is her behavior under low quality that is sensitive to the private information of D: The more accurate D's private information, the more likely it is that E gives a negative review. In this sense, E becomes more "honest" as D becomes more informed. The idea is that the positive review of E has to counteract the negative signal of D (recall that this is the only information set where D does not necessarily follow the review). When D's private signal becomes more robust, the positive review has to convey more information in order for E to avoid being ignored. In the limit where D knows the state for certain, E's messages are fully revealing. In the limit where D "knows nothing", E always gives a positive review regardless of the actual quality.

The change in the behavior of D is less straightforward. It is useful to highlight the following observation.

Remark 2. *In equilibrium, $\sigma^{1,0}$ is strictly increasing w.r.t. p. That is, when D's private signal is more accurate, she is more likely act against her own negative signal in favor of a positive review by E.*

It may seem counterintuitive at first glance that a better informed D is more likely to ignore her own information and act according to E's advice. Recall, however, that E uses true state ω to predict the private signal s of D, which helps her predict the reaction to her message m. The decision of E in the low quality state is determined by a trade-off between giving a negative review and being followed versus giving a positive review and potentially selling the product. But the latter comes at the risk of being ignored, which E suffers additionally from due to her influence motivation. Now suppose D is very well informed herself and E observes low quality. Then E knows it is very likely that D's private opinion will be negative and the risk from being ignored is too large. In other words, E is too inclined to recommend the action that she thinks D would take on her own, just so her recommendation is in line with D's action. Therefore, her fear of being ignored needs to be mitigated by a more compliant response by D.

To summarize the intuition; when D is better informed, E has a better estimate of her private opinion. This tempts E to stop "telling D what to do" and start "saying what she thinks D will do anyway". In this sense, when the decision maker is well informed, the causality between advice and action reverses itself: The expert ceases to be a leader and becomes a follower.

5.2 Decision Maker's Gain from Communication

Next I look at the welfare of D in the informative equilibrium. In particular: Is she better off by hearing E's advice, and if so, by how much?[18] How is D's utility affected the quality of her private information?

The natural benchmark for D's no-communication utility is the level that she receives when she plays $\sigma_D = (1, 0, 1, 0)$. That is, she plays $a = 1$ if and only if $s = 1$, following only her private signal. Henceforth, I refer to this as the autarky case. Denoting the autarky payoff by $\underline{u}_D$, we have $\underline{u}_D = \frac{2p-1}{2}$. Denoting D's ex-ante expected utility in the informative equilibrium of the communication game by $U_D(\sigma^*)$, we have $U_D(\sigma^*) = \frac{2p-1}{2p}$. Figure 5 plots these two payoffs against each other as a function of private signal accuracy p.

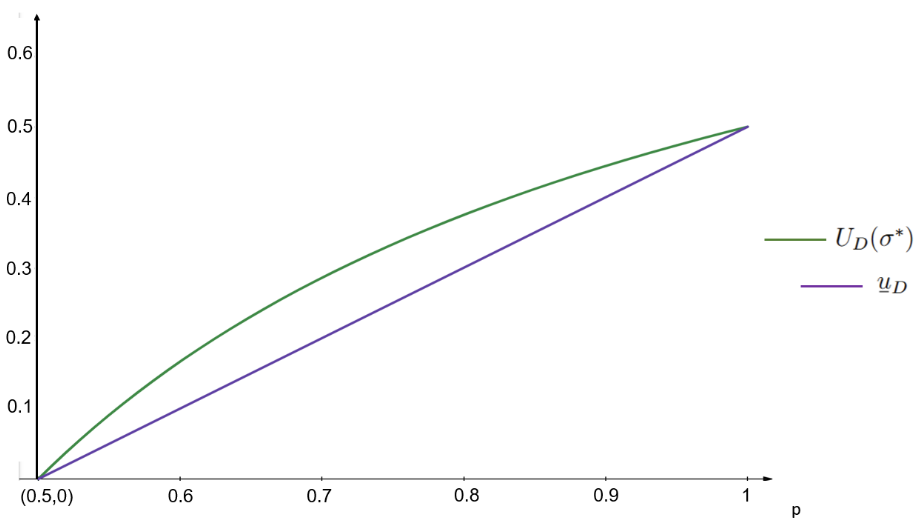

Figure 5: Utility of D as a function of p, given $\beta = 2/3$. The purple and green lines represent the autarky and informative equilibrium payoffs respectively.

Let us define D's *gain from communication* as $U_D(\sigma^*) - \underline{u}_D$. The main observations about the behavior of this gain with respect to the private signal accuracy can be summarized as follows.

[18]This measure is understudied in both the cheap talk and Bayesian Persuasion literature. The main concern in these strands of literature is the welfare of the sender (here E). This manifests itself further in a common selection criterion; the "sender-preferred equilibrium". Note that in contrast, the informative equilibrium in my model is unique.

Remark 3. *D's gain from communication is non-monotonic w.r.t. p. There exists a unique* $\hat{p} \in (0.5, 1)$ *that maximizes it. At both the fully uninformed* ($p \to 0.5$) *and perfectly informed* ($p \to 1$) *limits, there is no gain from communication.*

Therefore, E's advice is most valuable when D is neither too informed or too ignorant. In both extremes, the value of communication disappears. In the extreme ($p \to 0.5$) where D has no private information, the strategy of E converges to giving a positive review regardless of the state, as can be seen in Figure 4. Advice no longer contains any information, because E is too tempted to induce D to buy and D has no alternative source of information to counteract her advice. In the extreme ($p \to 1$) where D knows the state for certain, E converges to perfect honesty. However, the need for additional information disappears for D and there is nothing to be gained from listening to advice.

In short, a D who is too uninformed is too open to "manipulation" for E to be truthful. A D who is too well informed does not need E to take the right action in the first place. There is a unique interior level of private information for D that makes communication the most useful.

5.3 Private Signal: Interpretation and Takeaways

So far, we have been relatively agnostic about the source of D's private signal, but it can be relevant to the meaning of the above results. A straightforward interpretation is that s is D's innate knowledge about the topic (e.g., has experience with the type of product in question). In that case, for instance, we can say that a individuals that are completely in the dark about the subject, as well as those who are experts in their own right, need not consult with "better experts".

Another possibility is that D has done some research of her own regarding this specific decision. This would be best captured by a private signal whose precision is costly to increase. In that case, the interior optimum for p tells us that if D is going to consult with an expert later on, she should neither do too much or too little prior research. This is especially true if the amount (but not the outcome) of prior research is known by the expert.

We can also consider s as the posterior after having consulted with another -less informed-expert before this interaction.[19] This interpretation provides insights about the complementarity between the informedness of different experts: Communication is useless if the previous expert is known to have been almost clueless or almost exactly as knowledgeable as the current one. The maximum gain lies somewhere in between. Therefore, the informedness of experts are complements when they are very different (i.e., low p), and substitutes when they are close to each other (i.e., high p).

Finally, we look at the effect of having private signals as opposed to them being observed by E.[20]

Remark 4. *Suppose that E observes the realization of s prior to sending her message. Then in any equilibrium, D receives utility $\underline{u}_D$.*

If D's opinion is common knowledge, then there is no equilibrium where m is informative enough to affect D's action. If E can condition her messages on D's signal, she cannot induce different actions under difference states in an incentive compatible way. Therefore, communication is useless and D is effectively in autarky. This underlines the importance of having uncertainty on the side of E. In short, it means that if D wants to benefit from consulting with an expert, then she should keep her opinion "a secret". This observation can have implications in many areas. For instance, it serves as an argument in favor of data privacy that takes place entirely in the framework of a single, one-shot interaction.

6 Extension: Influence versus Commitment with Higher Dimensions

Here, I extend the main 2×2 model to allow for an arbitrary (finite) number of states and D actions. In this generalized setting, I revisit the comparison between the influence-motivated game and the purely-biased commitment problem.

[19] This projection is made without taking into account the strategic concerns of the "first-expert". One way to think of this is as a "subgame" that starts at the interaction with the last expert. Another way is to suppose that the earlier expert was not aware that the interaction we analyze would happen. Finally, the previous expert may have been "benevolent", in the sense of fully revealing her (noisy) information.

[20] This is the same model as D not receiving any signals but having a known prior $\mu_0 \in \{1-p, p\}$

Specifically, I compare two solutions: One is the optimal "simple" (i.e., binary) commitment strategy of an expert E facing an uninformed D. The other is the sender-preferred equilibrium of the influence-motivated model in which D is privately informed. I provide sufficient conditions under which the unique optimal message structure in the commitment problem emerges as the limit of the sender-preferred equilibrium, as D's private information vanishes. In other words, an expert who can ex-ante commit, but is constrained to a "recommend or not" technology, behaves identically to an influence-motivated expert who is free to use as many messages as she wishes but cannot commit.

The intuition underlying this limiting equivalence is the same as in the 2×2 model, as I discuss in detail below. Note also that neither the conditions of Theorem 3 nor the "binary commitment message" constraint are *necessary* for a behavioral equivalence between the two models. A full characterization includes many other specifications where the equivalence holds. However, the discussion below furthers the intuition and the scope of the parallel between the two types of experts.

Before stating the result, I introduce the modified setting and define the strategies of interest.

6.1 Setting

Define set of states $\omega \in \Omega$, actions $a \in A$, and messages $m \in M$ with $\Omega = M = A = \{1, ..., N\}$. Suppose N is an odd natural number, and priors over Ω are uniform.[21] As in the two-by-two specification, DM observes a binary private signal $s \in \{0, 1\}$ about the state. Suppose the probability of receiving the high signal $s = 1$ under state ω is given by $p(s = 1|\omega) = \eta\pi_\omega + (1-\eta)0.5$ where $\pi_\omega, \eta \in (0, 1)$, and π_ω is strictly increasing in ω (so signal s satisfies the monotone likelihood ratio property). That is, s is informative with probability η and noise with probability $1 - \eta$. This structure allows us to vary the informativeness of s via η.

[21]The odd number of states and the equal priors assumptions are not essential to the result below. They are made to simplify the discussion by ruling out the non-generic case of indifference under the prior.

6.2 Influence Motivated Game

Payoffs: The utility of the expert is given by $U_E(a,m) = \beta u(a) - (a-m)^2$ with some strictly increasing function $u : A \to [0,1]$ and $\beta > 0$. That is, actions are ordered and their magnitude are of meaning to both agents (e.g., how much of a given product to purchase). Expert E is motivated by two distinct factors: The action is high as possible and the action of D is as close to her recommendation as possible. The highest and the lowest payoffs are normalized to $u(1) = 0$ and $u(N) = 1$. The payoff function of D is the standard quadratic loss function where she wants to match the state: $U_D = -(a-\omega)^2$

Object of interest: Denote a mixed strategy of the expert by $\sigma_E : \Omega \to \Delta M$, and a mixed strategy of the decision maker by $\sigma_D : M \times \{0,1\} \to \Delta A$ as a mapping from a message and private signal pair to an action. As an extension of the class of strategies we considered in the two-by-two case, call an expert strategy "non-pooling" if it does not send the same message for certain under every state. That is, for all $m \in \{1, ..., N\}$, there exists some $\omega \in \{1, ..., N\}$ such that $p(m|\omega, \sigma_E) < 1$.[22] Denote the set of all such strategies by Σ_E^{NP}.

Finally, we define the set of equilibria that are sender-preferred among those that are non-pooling. Since the main result of this section concerns the limit as the DM becomes uninformed, let $\Sigma^*(\eta)$ be the set of all equilibrium strategy profiles $\sigma = (\sigma_E, \sigma_D)$ as a correspondence from η. We can then write the set of sender-preferred non-pooling equilibria as $\bar{\Sigma}(\eta) := \text{argmax}_{\sigma \in \{\sigma' \in \Sigma^*(\eta) | \sigma'_E \in \Sigma_E^{NP}\}} \mathbb{E}_\omega \mathbb{E}_\sigma [U_E(a,m)]$.

6.3 Uninformed Commitment Problem

Now consider three modifications to the influence motivated game above: (i) The DM does not receive a private signal (i.e., suppose $\eta = 0$), (ii) Expert can ex-ante commit to a message strategy and (iii) Her objective function is simply $u(a)$, where the function u is as defined in the influence motivated case.

[22] As before, the term "non-pooling" here is used in a non-standard way as a shorthand to rule out strategies that send the same message for certain under all states. The result below allows for mixed pooling strategies.

Expert's problem: We can restrict the set of messages available to the expert to M without loss of generality.[23] The expert's problem is to choose $\sigma_{BP} : \Omega \to \Delta(M)$ maximizing $\mathbb{E}_{\sigma_{BP}}[u(a)]$, where the expectation is taken over states, messages, and the DM's subsequent sender-preferred best-response action a, which maximizes her utility U_D given the posterior belief induced by each message.[24]

Object of interest: Here the goal is to compare the equilibrium of the influence motivated model to the optimal commitment strategy among the simple "recommend or not" information structures. First, let $supp(\sigma_{BP})$ denote the set of messages $m \in M$ such that $\sigma_{BP}(m) > 0$ for some $\omega \in \Omega$. That is, $supp(\sigma_{BP})$ is the set of messages that σ_{BP} sends with positive probability under some state. Call a message strategy σ^*_{BP} *"binary-constrained optimal"* if $\sigma^*_{BP} \in \arg\max_{\{\sigma_{BP}:|supp(\sigma_{BP})|\leq 2\}} \mathbb{E}_{\sigma_{BP}}[u(a)]$

Second, recall that unlike the equilibria of the influence motivated game, the labels of these strategies are arbitrary. I therefore group them into **equivalence classes** based on relabeling. The *equivalence class* of a strategy σ_{BP}, denoted $[\sigma_{BP}]$, is the set of all strategies that are relabelings of it. Formally, $\sigma'_{BP} \in [\sigma_{BP}]$ if there exists a permutation $\pi : M \to M$ such that for all $\omega \in \Omega$ and $m \in M$:

$$\sigma_{BP}(m|\omega) = \sigma'_{BP}(\pi(m)|\omega)$$

where $\sigma_{BP}(m|\omega)$ is the probability of sending message m in state ω. The set $[\sigma^*_{BP}]$, consisting of all relabelings of a constrained optimal strategy, serves as the benchmark in the comparison to the influence-motivated equilibrium.

6.4 Result

To capture the convexity of function u, denote by $\Delta^2_u(a) := \frac{u(a+1)-u(a)}{u(a)-u(a-1)}$ its percentage second difference (i.e., growth rate of the marginal utility) at $a \in \{2, .., N-1\}$. The following result

[23] As discussed in the 2×2 case, any outcome under an arbitrary signal policy can also be achieved with a policy where the messages are direct recommendations for the DM's actions (Proposition 1 of Kamenica and Gentzkow, 2011). Therefore, we can restrict attention to message spaces no larger than the state space.

[24] Note that since the problem is formulated directly in terms of messages, we do not need to state the Bayes-Plausibility constraint. It is implicitly satisfied by the definition of a message structure.

relates the sender preferred equilibrium to the optimal commitment solution.

Theorem 3. *There exist* $\bar{\beta} > 0$ *and* $\bar{\Delta} > 1$ *such that the following hold for* $\beta < \bar{\beta}$ *and* $\min_{a \in \{2,\ldots,N-1\}} \Delta_u^2(a) > \bar{\Delta}$*:*

(i) The commitment problem with the uninformed DM yields a unique equivalence class of message strategies that are constrained-optimal, denoted by $[\sigma^*_{BP}]$*.*

(ii) For η *sufficiently small, all equilibria in* $\bar{\Sigma}(\eta)$ *include the same unique expert strategy, denoted* $\bar{\sigma}_E(\eta)$*. For this strategy, we have*

$$\lim_{\eta \to 0} \bar{\sigma}_E(\eta) \in [\sigma^*_{BP}]$$

That is, any constrained-optimal commitment strategy is a relabeling of the pointwise limit of $\bar{\sigma}_E(\eta)$*.*

To restate the result: If (*i*) the influence motivation is sufficiently large, and (*ii*) the common "bias" component u of the E utility is sufficiently convex in action a, then there is a unique binary message structure that is constrained-optimal in the commitment problem with uninformed D. This message structure is identical to the unique sender preferred equilibrium of the influence motivated model in the corresponding limit (i.e., as D becomes uninformed). Moreover, any specific constrained-optimal commitment strategy is merely this influence motivated equilibrium strategy with the names of the messages rearranged.

The basic idea in plain words is therefore: In the limit, a highly influence motivated strategic communicator (i.e., without commitment) mimics the thumbs up/thumbs down behavior of an ambitious information designer (i.e., with commitment).

The optimal commitment strategy in question is a binary *interval partition* of Ω. That is, there is a $\bar{\omega}$ such that E sends some message $m \in M$ for certain if $\omega < \bar{\omega}$ and another message $m' \in M/\{m\}$ for certain if $\omega \geq \bar{\omega}$. The sender preferred equilibrium of the influence motivated game has precisely the same structure, except at one state: At the cutoff state $\bar{\omega}$, E mixes between the high and low messages. As $\eta \to 0$, however, the mixing vanishes and the strategy of E converges to sending the high message for certain in the cutoff state. Therefore, the structures become identical in the limit.

The intuition is similar to the 2×2 case. Since u is convex, the information designer faces a bad status quo and is "ambitious": She "targets" a particular action that is higher than the one under the prior, and aims to maximize the probability of the message that induces this action. Therefore, she sends the high message precisely so often that D is indifferent between the target action and the one below it. The low message induces a "dump posterior" and the payoff upon receiving it is secondary.

The influence motivated E prefers the same "target action". To maintain the incentive compatibility of sending the low message, however, the high message must be ignored with positive probability, which in turn requires indifference. For the informed receiver, the indifference point has to be achieved for the "pessimistic" type of D (i.e., $s = 0$). This ensures that the optimistic type follows the high message. As η decreases, the indifference points of the two types of D converge to one another. In the limit, the required indifference posterior mean, that is meant to be induced as often as possible, matches the uninformed commitment case.

Hence, in the limit as $\eta \to 0$, the influence-motivated equilibrium strategy of E coincides with the binary commitment outcome

Finally, this generalized setting reinforces our observation on the semantics of message labels. As in the main model, the names of the constrained-optimal commitment messages are arbitrary. In contrast, the two equilibrium path messages of the influence motivated game *must* take the name of the action closest to their induced posterior means (with ties broken in favor of the higher action). For instance, if the "target action" is the highest action N, then E sends $m = N$ in the two highest states, and $\frac{N-1}{2}$ (the posterior mean conditional on the low message) for all $\omega < N - 1$.[25]

Hence, the comparison of the two models suggests the following interpretation: if "thumbs up" or "thumbs down" are communicated through action recommendations, then the expert chooses their labels to reflect the expected state conditional on the message. Conversely, when an action is recommended by name, the decision maker should interpret it as representing a distribution of states whose mean equals that recommendation.

[25]This exact structure is due to uniform priors, but the observation that messages must carry the name of the induced posterior mean maintains for all priors.

7 Concluding Remarks

This paper accounts for a natural concern of an advisor in strategic communication—being influential. I show that this concern can reconcile the behavior predicted by Bayesian persuasion with a key institutional feature commonly observed in practice: the inability to commit. By factoring in experts' intrinsic desire to be followed, we can forgo the commitment assumption, which otherwise limits the applicability of the information design framework. In other words, through influence motivation, the predictions of information design remain highly relevant even when advisors lack commitment power.

The influence motivated model itself yields important insights about the nature of communication. The conventional view that recommendations shape actions holds only when the listener does not possess sufficient information of her own. Once she "knows what she is doing" to a reasonable extent, it is her action—as anticipated by her advisors—that dictates what is recommended to her.

The analysis offers policy-relevant guidance to decision makers, especially regarding the amount and "privacy" of private information. A decision maker should know neither too much nor too little beforehand if she wants to make the most of the interaction: if she knows too little, she is easy to manipulate; if she knows too much, the advisor merely tries to tell her something she would agree with. Knowing the optimal amount of private information can help the decision maker decide whom to consult beforehand, or how much research she should do on her own. Finally, and perhaps most significantly, listening to advice is useless for the decision maker if the expert is aware of what she thinks. Therefore, she should not share her opinions with the advisor if she wants any beneficial advice. This observation serves as a strong case for safeguarding the privacy of personal history and preference data.

Appendix A: Main Model

Theorem 1. *If $\beta < \frac{p}{1-p}$, then there is a unique equilibrium that satisfies $\sigma_E^* \notin \{(1,1),(0,0)\}$. The strategies are as follows.*

$$\sigma_E^* = \left(1, \frac{1-p}{p}\right) \quad \wedge \quad \sigma_D^* = \left(1, \frac{p(1+\beta)-\beta}{p(1+\beta)}, 0, 0\right)$$

Otherwise, $\sigma_E^* \in \{(1,1),(0,0)\}$ *in any equilibrium.*

Proof. We restrict attention to "non-pooling" candidate strategies by E. That is $\sigma_E \notin \{(0,0),(1,1)\}$. Under strategy $\sigma_E = (\sigma^1, \sigma^0)$, denote by $\mu(m,s)$ the posterior belief of D upon observing signal s and message m. We have:

$$\mu(1,s) = \frac{p(s|\omega=1)\sigma^1}{p(s|\omega=1)\sigma^1 + p(s|\omega=0)\sigma^0}$$
$$\mu(0,s) = \frac{p(s|\omega=1)(1-\sigma^1)}{p(s|\omega=1)(1-\sigma^1) + p(s|\omega=0)(1-\sigma^0)}$$

With this notation, we have that $a=1$ is a best response to (m,s) is if and only if $\mu(m,s) \geq 0.5$ Since $\mu(m,0) \geq 0.5$ implies $\mu(1,m) > 0.5$ we can make the following observation on the equilibrium behavior of D

Observation 1. *For any* $m \in \{0,1\}$*, if* σ_D *is a best response, then* $\sigma^{m,0} > 0$ *implies* $\sigma^{m,1} = 1$.

With this in mind, we turn to E. Given σ_D, denote by $U_E(m|\omega)$ the expected utility of E from sending message m upon observing state ω. These can be written as:

$$U(1|\omega) = (1+\beta)\sum_{s\in\{0,1\}} p(s|\omega)\sigma^{1,s}$$
$$U(0|\omega) = \sum_{s\in\{0,1\}} p(s|\omega)(\beta\sigma^{0,s} + 1 - \sigma^{0,s})$$

Next, define the gain function $g(\omega) := U(1|\omega) - U(0|\omega) \geq 0$. We can write this as follows.

$$g(\omega) = \sum_{s\in\{0,1\}} p(s|\omega)\left((1+\beta)\sigma^{1,s} + (1-\beta)\sigma^{0,s} - 1\right)$$

which leads to the second observation.

Observation 2. *For any* $\omega \in \{0,1\}$, $m=1$ *is a best response for E if and only if* $g(\omega) \geq 0$.

Now we can use these observations to derive the equilibrium by eliminating "non-pooling" candidates in four steps.

Step 1. *In equilibrium,* $m=1$ *cannot perfectly reveal either state. That is, we must have* $(\sigma^1, \sigma^0) >> (0,0)$.

To see this by contradiction, first assume $\sigma^1 > \sigma^0 = 0$. This, for the best response of D implies $\sigma^{m,s} = 1$ for any $m \in \{0, 1\}$. We then have

$$g(0) = \sum_{s \in \{0,1\}} p(s|0)(\beta + (1 - \beta)\sigma^{0,s}) > 0$$

which means E has incentive to deviate to $m = 1$ under $\omega = 0$. Assuming the complementary case, we have $\sigma^0 > \sigma^1 = 0$ which implies $\sigma^{1,0} = \sigma^{0,0} = 0$. This leads to

$$g(0) = \sum_{s \in \{0,1\}} p(s|0)((1 - \beta)\sigma^{0,s} - 1) < 0$$

so E has incentive to deviate to $m = 0$ under $\omega = 0$. This concludes the statement of Step 1.

Step 2. *There is no σ_D that is a best response to a non-pooling σ_E and makes E indifferent under both states. Therefore, we must have $(\sigma^1, \sigma^0) \notin (0, 1)^2$ in equilibrium.*

To see this, first note that indifference under both states requires $g(1) = g(0) = 0$. The first equation ($g(1) = g(0)$) is equivalent to the following.

$$(1 + \beta)\sigma^{1,1} + (1 - \beta)\sigma^{0,1} - 1 = (1 + \beta)\sigma^{1,0} + (1 - \beta)\sigma^{0,0} - 1$$

and since both gains have to be equal to zero, the condition for indifference in both states is equivalent to

$$(1 + \beta)\sigma^{1,1} + (1 - \beta)\sigma^{0,1} = 1 \tag{1}$$

$$(1 + \beta)\sigma^{1,0} + (1 - \beta)\sigma^{0,0} = 1 \tag{2}$$

However, this is not possible: First consider the case where $\sigma^{1,0} = 0$. This, from condition (2), should imply $\sigma^{0,0} = \frac{1}{1-\beta}$. This is impossible for any mixed strategy since $\frac{1}{1-\beta} < 0$ if $\beta > 1$, and $\frac{1}{1-\beta} < 0$ otherwise.

Second, take the complementary case of $\sigma^{1,0} > 0$. This, from Observation 1 above, must imply $\sigma^{1,1} = 1$. However, this means from condition (1) that $\sigma^{0,1} = \frac{\beta}{1-\beta}$. This is impossible because $\frac{\beta}{1-\beta} < 0$ if $\beta < 1$ and $\frac{\beta}{1-\beta} > 1$ otherwise. This exhausts the possible strategies of D that makes E indifferent under both states and therefore completes Step 2.

Step 3. *In equilibrium, $m = 0$ cannot perfectly reveal $\omega = 1$. Therefore, we must have $\sigma^0 < 1$.*

To see this by contradiction, take strategy $(\sigma^1, 1)$ with $\sigma^1 \in (0, 1)$. This form is the only remaining one that satisfies Steps 1 and 2 and violates Step 3.

Such σ_E implies $\sigma^{0,1} = \sigma^{0,0} = 1$ and $\sigma^{1,0} = 0$ for the best response of D. Then for E's expected utility, we have the following.

$$U(0|1) = U(0|0) = \beta$$
$$U(1|1) = (1+\beta)p\sigma^{1,1}$$
$$U(1|0) = (1+\beta)(1-p)\sigma^{1,1}$$

So for the strategy $(\sigma^1, 1)$ to be optimal, we then need the two following conditions

$$U(1|0) \geq U(0|0) = \beta \tag{3}$$
$$U(1|1) = U(0|1) = \beta \tag{4}$$

However, this leads to a contradiction since $U(1|0) = \beta$ implies $\sigma^{1,1} > 0$ from Observation 1. In turn, $\sigma^{1,1} > 0$ implies $U(1|1) > U(1|0)$. Therefore, conditions (3) and (4) are contradictory. Therefore, if E is indifferent under $\omega = 0$, then she must strictly prefer $m = 1$ under $\omega = 1$. This completes the statement of Step 3.

Step 4. *The unique non-pooling equilibrium of the game is given by the strategies described in Theorem 1 if $\beta \leq \frac{p}{1-p}$. No such equilibrium exists otherwise.*

The only remaining non-pooling candidates from Steps 1 through 3 is of the form $(\sigma^1, \sigma^0) = (1, \sigma^0)$ with $\sigma^0 \in (0, 1)$. This means $m = 0$ is a perfect signal of $\omega = 0$. This means $\mu(0, 0) = \mu(0, 1) = 0$ and therefore $\sigma^{0,0} = \sigma^{0,1} = 0$. Furthermore,it means $\mu(1, 1) = \frac{p}{p+(1-p)\sigma^0} > 0.5$ and therefore $\sigma^{1,1} = 1$.

For a σ_E of this for to be a best response, E must be indifferent under $\omega = 0$. Plugging in the above observations on σ_D, we can write the indifference condition as follows and solve for $\sigma^{1,0}$.

$$g(0) = (1-p)\beta + p\left((1+\beta)\sigma^{1,0} - 1\right) = 0$$
$$\Leftrightarrow \sigma^{1,0} = \frac{p(1+\beta) - \beta}{p(1+\beta)}$$

For this to be a best response by D, she must be indifferent in the information set $m = 1$ with $s = 0$. That is,

$$\mu(1,0) = \frac{(1-p)\sigma^1}{(1-p)\sigma^1 + p\sigma^0} = 0.5$$
$$\Leftrightarrow \sigma^0 = \frac{1-p}{p}$$

which completes the strategy profile of the Theorem. Note that an equilibrium of this type only exists if $\beta \leq \frac{p}{1-p}$ as this is necessary for $\sigma^{1,0} \geq 0$ in the indifference condition of E under $\omega = 0$.

□

Appendix B: Multi-State Extension (Section 6)

Theorem 3. *There exist $\bar{\beta} > 0$ and $\bar{\Delta} > 1$ such that the following hold for $\beta < \bar{\beta}$ and $\min_{a \in \{2,\ldots,N-1\}} \Delta_u^2(a) > \bar{\Delta}$:*

(i) The commitment problem with the uninformed DM yields a unique equivalence class of message strategies that are constrained-optimal, denoted by $[\sigma^*_{BP}]$.

(ii) For η sufficiently small, all equilibria in $\bar{\Sigma}(\eta)$ include the same unique expert strategy, denoted $\bar{\sigma}_E(\eta)$. For this strategy, we have

$$\lim_{\eta \to 0} \bar{\sigma}_E(\eta) \in [\sigma^*_{BP}]$$

That is, any constrained-optimal commitment strategy is a relabeling of the pointwise limit of $\bar{\sigma}_E(\eta)$.

Proof. The proof first derives the sender-preferred non-pooling equilibrium of the influence motivated model for sufficiently small η, and establishes the expert strategy in the limit.

Second, it derives the optimal signal strategy of the commitment problem and shows the equivalence.

For brevity during the rest of the proof, denote $p_\omega := p(s = 1|\omega) = \eta\pi_\omega + (1 - \eta)0.5$. Recall that p_ω is strictly increasing in ω.

Influence Motivated Equilibrium

With some abuse of notation, denote by $\sigma_\omega(m)$ the probability that the expert sends message m under state ω. Similarly, let $\sigma_{m,s}(a)$ denote the probability that D plays action a after observing message m and private signal s. To describe the sender payoffs under strategy profile (σ_E, σ_D), first let $v(m|s) := \mathbb{E}_{a\sim\sigma_{m,s}}(U_E(a, m))$ denote the expected (over mixed strategy of D) expert utility when she sends message m and the private signal is s. Then we can define the expected expert utility from sending m under state ω as follows.

$$v_\omega(m) := p_\omega v(m|1) + (1 - p_\omega)v(m|0)$$

Finally, denote by $\tilde{M}_\sigma := \{m \in \{1, ..., N\}|\exists\omega : \sigma_\omega(m) > 0\}$ as the set of on path messages under strategy σ_E. The condition for optimality (i.e., "incentive constraint") for the expert is that for any m with $\sigma_\omega(m) > 0$, we must have $m \in argmax_{m'\in\{1,...,N\}} v_\omega(m')$.

Using this notation, we can now identify the sender-preferred non-pooling equilibrium strategy of E. For sufficiently small β and η, I first establish the candidate equilibrium and then show that it is indeed the sender preferred among non-pooling equilibria.

Candidate Equilibrium: Consider strategy profile with $\bar{\sigma}_E$ given by $\bar{\sigma}_\omega(\frac{N-1}{2}) = 1$ for all $\omega \leq N-2$, $\bar{\sigma}_N(N) = 1$, and $\bar{\sigma}_{N-1}(N) = 1 - \bar{\sigma}_{N-1}(\frac{N-1}{2}) = q_E$ with $q_E = \frac{1-p_N}{1-p_{N-1}}$. That is, E's strategy is a binary partition of the state space, where she sends message N if the state is N, message $\frac{N-1}{2}$ if the state is smaller than $N-1$, and in state $N-1$, she mixes between the two messages according to the likelihood of receiving the low signal under state N relative to $N-1$.

Optimally, D chooses the action that is the closest to her posterior mean upon observing m and s. This is given by:

$$\tilde{\mu}(m, s) := \sum_{\omega=1}^{N} \omega \frac{p(s|\omega)\sigma_\omega(m)}{\sum_{\omega'=1}^{N} p(s|\omega')\sigma_{\omega'}(m)}$$

For E's strategy $\bar{\sigma}_E$, and η sufficiently small, we have the following posterior means following the on-path messages $\tilde{M}_{\bar{\sigma}} = \{\frac{N-1}{2}, N\}$ and private signal s:

$$\tilde{\mu}\left(\frac{N-1}{2}, s\right) \in \left(\frac{N-2}{2}, \frac{N}{2}\right) \quad \text{for any } s \in \{0,1\}$$
$$\tilde{\mu}(N,0) = N - \frac{1}{2}$$
$$\tilde{\mu}(N,1) > N - \frac{1}{2}$$

The first case is confirmed by $\lim_{\eta \to 0} q = 1$ and $\lim_{\eta \to 0} p(0|\omega) = \lim_{\eta \to 0} p(1|\omega) = 0.5$ for all ω, which means $\tilde{\mu}(\frac{N-1}{2}, s)$ converges to $\frac{N-1}{2}$ for any s. Therefore, if D observes $m = \frac{N-1}{2}$, it is strictly preferred to play $a = m$. The same is true if she observes $m = N$ with $s = 1$. When she observes $m = N$ along with $s = 0$, then she is indifferent between $a = N - 1$ and $a = N$.

Suppose now that D plays strategy $\bar{\sigma_D}$ such that: On path (i.e., $m \in \{\frac{N-1}{2}, N\}$) she plays $a = m$ for certain at every information set except $m = N$ and $s = 0$. If $m = N$ and $s = 0$, she mixes between actions N and $N - 1$ with $\sigma_{N,0}(N) = 1 - \sigma_{N,0}(N-1) = q_D$, where q_D is given by

$$q_D = \frac{\beta[u(\frac{N-1}{2}) - (p_{N-1}u(N) + (1 - p_{N-1})u(N-1))] + 1 - p_{N-1}}{(1 - p_{N-1})[\beta(u(N) - u(N-1)) + 1]} \tag{A}$$

Note that for sufficiently small β, we have $q_D > 0$. Furthermore, $\lim_{\beta \to 0} q_D = 1$. Therefore, this mixed strategy indeed exists for all sufficiently small β.

This value of q_D solves $v_{N-1}(\frac{N-1}{2}) = v_{N-1}(N)$, so it indeed makes E indifferent between $m = \frac{N-1}{2}$ and $m = N$ in state $N - 1$. For the incentive of E in other states given $\bar{\sigma}_D$, note that under $v_\omega(\frac{N-1}{2}) = \beta u(\frac{N-1}{2})$ for all ω. For the payoff from playing $m = N$, we have

$$v_\omega(N) = p_\omega \beta u(N) + (1 - p_\omega)(q_D \beta u(N) + (1 - q_D)(\beta u(N-1) - 1))$$

which is equal to $v_\omega(\frac{N-1}{2})$ for $\omega = N - 1$. Since (i) p_ω is strictly increasing in ω and (ii) $\beta u(N)$ is the highest feasible expert payoff in the game, $v_\omega(N)$ is strictly increasing in ω. Therefore, the indifference in state $N - 1$ implies that $m = \frac{N-1}{2}$ is strictly preferred over $m = N$ for all $\omega < N - 1$ and $m = N$ is strictly preferred over $m = \frac{N-1}{2}$ for $\omega = N$. Hence, the equilibrium path messages of $\bar{\sigma}_E$ are indeed incentive compatible given $\bar{\sigma}_D$.

Finally, to maintain incentive compatibility with respect to off-path messages, suppose that $\tilde{\mu}(m.s) = 1$ for all $m \notin \{\frac{N-1}{2}, N\}$. The resulting best response is $\bar{\sigma}_D(m, s) = 1$ for all

s, giving E a payoff of $\beta u(1) - (m-1)^2 = -(m-1)^2 < \min\{v_\omega(\frac{N-1}{2}), v_\omega(N)\}$ for all ω. Therefore, E has no incentive to deviate to any of the off-path messages.

Hence, for sufficiently small β and η, $\bar{\sigma}$ as defined above is an equilibrium of the game. We can write the ex-ante expert payoff from this equilibrium as follows:

$$\mathbb{E}_{\bar{\sigma}} U_E = \frac{N-1}{N}\beta u\Big(\frac{N-1}{2}\Big) + \frac{1}{N}\Big[(p_N + (1-p_N)q_D)\beta u(N) + (1-p_N)(1-q_D)(\beta u(N-1) - 1)\Big] \tag{B}$$

Sender-Preferedness: Now we fix a β sufficiently small such that $\bar{\sigma}$ is an equilibrium, and show that it yields the highest ex-ante utility for E among all non-pooling equilibria for sufficiently small η. This follows 3 steps of establishing properties that a candidate profile should satisfy.

Step 1: Never "ignore for certain": In other words, any candidate profile must satisfy $\sigma_{m,s}(m) > 0$ for all $m \in \tilde{M}$, and all $s \in \{0,1\}$.

To see this, take any profile σ with $\sigma_{m',s'}(m') = 0$ for some (m', s') with $m' \in \tilde{M}$. Let ω' be a state under which m' is sent with positive probability. First note that if σ_E is optimal given σ_D, then the expected payoff of E under state ω' must be $v_{\omega'}(m')$. Recall that we can write this as.

$$v_{\omega'}(m') = p(s'|\omega')\mathbb{E}_{a\sim\sigma_{m',s'}}(\beta u(a) - (m'-a)^2) + p(s''|\omega')v(m'|s'')$$

where s'' is the other signal $s'' \in \{0,1\}/\{s'\}$. Since $\sigma_{m',s'}(m') = 0$, we know that the first additive term (payoff from (m', s')) is bounded above by -1 as β goes to 0. Also note that by construction of the utility, we have $\lim_{\beta\to 0} v(m|s) \leq 0$ for any m, s. Applying this to m' with s'', we see that $\lim_{\beta\to 0} v(m'|s'') \leq 0$ and therefore $\lim_{\beta\to 0} v_{\omega'}(m') \leq -p(s'|\omega')$.

Recall that $\bar{\sigma}$ is the equilibrium profile we established above. Now we can compare its ex-ante expert utility to strategy profile σ for low β as follows:

$$\begin{aligned}\lim_{\beta\to 0} \mathbb{E}_{\sigma} U_E &= \frac{v_{\omega'}(m')}{N} + \frac{1}{N}\sum_{\omega\neq\omega'} \mathbb{E}_{m\sim\sigma_E} v_\omega(m) \\ &\leq \frac{\lim_{\beta\to 0} v_{\omega'}(m')}{N} \leq -\frac{p(s'|\omega')}{N} < 0\end{aligned}$$

where the first inequality is from the property $\lim_{\beta\to 0} v(m|s) \leq 0$ discussed above.

Looking at Equation (B), on the other hand, we see that for the candidate equilibrium $\bar{\sigma}$,we have $\lim_{\beta\to 0} \mathbb{E}_{\bar{\sigma}} U_E = 0$. This follows because $\lim_{\beta\to 0} q_D = 1$.

Finally note that the expected E payoffs from both strategy profiles are continuous in β. Hence, for sufficiently small β, equilibrium $\bar{\sigma}$ is better for E than any profile σ with some $m \in \tilde{M}$ and s such that $\sigma_{m,s}(m) = 0$.

Step 2: Sender-preferred equilibrium is an interval partition: That is, the message structure should satisfy the following: For all $m \in \tilde{M}$, there exist two states $\underline{\omega}_m \leq \bar{\omega}_m$ such that: (i) $\sigma_\omega(m) = 0$ for all $\omega \notin \{\underline{\omega}_m, ..., \bar{\omega}_m\}$, (ii) $\sigma_\omega(m) = 1$ for all $\omega \in \{\underline{\omega}_m + 1, ..., \bar{\omega}_m - 1\}$, and (iii) $|\{\underline{\omega}_m, ..., \bar{\omega}_m\} \cap \{\underline{\omega}_{m'}, ..., \bar{\omega}_{m'}\}| \leq 1$ for all $m' \neq m$. Furthermore, the partition is *ordered* in the sense that $m > m'$ implies $\underline{\omega}_m \geq \bar{\omega}_{m'}$.

To show this, we need to derive a necessary condition on the structure of σ_D in candidate strategy profiles: In any candidate equilibrium, for all $m \in \tilde{M}/min\{\tilde{M}\}$, we have $\sigma_{m,s}(m) \in (0,1)$ for some $s \in \{0,1\}$ and $\sigma_{m,s'}(m) = 1$ for $s' \neq s$. That is, for any on path message but the lowest, D must follow for certain under one private signal, and with an interior probability under the other.

We have already established in Step 1 that $\sigma_{m,s}(m) > 0$. Additionally, we cannot have an equilibrium where there is a message m that is followed for certain and is not the smallest message on path. If a message is followed for certain under any signal s, it gives utility $\beta u(m)$ regardless of the state. Denoting the smallest message by $\underline{m} := min\tilde{M}$, the payoff from sending $\underline{m}$ under any state is bounded above by $\beta u(\underline{m}) < \beta u(m)$. This contradicts $\underline{m} \in \tilde{M}$: message $\underline{m}$ is strictly worse and should not be sent. Hence, for any higher on-path message, there should be at least one $s \in \{0,1\}$ where D ignores with positive probability.

Finally, $\sigma_{m,s}(m) \in (0,1)$ means that the posterior mean after observing m and s is either $m - 0.5$ or $m + 0.5$. As discussed above, the posterior mean from (m, s') converges to the one from (m,s) as $\eta \to 0$. Fixing the posterior after (m,s) to $\tilde{\mu}(m,s) \in \{m-0.5, m+0.5\}$ as we have shown to be required, we must have $\tilde{\mu}(m,s') \in (m-0.5, m+0.5)$ for η small, meaning $\sigma_{m,s'}(m) = 1$. Otherwise, the best response of D yields $\sigma_{m,s'}(m) = 0$, which contradicts the observation of Step 1.

Hence, in any candidate profile, for all $m \in \tilde{M}/min\{\tilde{M}\}$, we have $\sigma_{m,s}(m) \in (0,1)$ for some $s \in \{0,1\}$ and $\sigma_{m,s'}(m) = 1$ for $s' \neq s$.

With this in mind, let us turn to the incentive constraint of E for fixed σ_D that satisfies the above property. Recall that if m is optimal in state ω, then we have $m \in argmax_{m'} v_\omega(m')$. To show the partition property, define

$$\begin{aligned} v(m,p) &:= pv(m|1) + (1-p)v(m|0) \\ &= p(v(m|1) - v(m|0)) + v(m|0) \end{aligned}$$

For fixed σ_D, $v(m|s)$ is independent of p. Therefore, $v(m,p)$ is a linear function of p. Since ω only affects $v(m,p)$ through p, we can rewrite the optimality condition for sending m in state ω with fixed σ_D as $m \in argmax_{m'} v(m', p_\omega)$. Now note that

1. For all m, $v(m,p)$ is linear in p.

2. For all $m \neq m'$, there exists p with $v(m,p) \neq v(m',p)$.

The second observation follows from the structure of σ_D we derived above. To see this, suppose without loss that $m > m'$. This means $\beta u(m) \in \{v(m|1), v(m|0)\}$ (since m must be followed for certain after some signal). By the utility function of E, this means

$$\max\{v(m|1), v(m|0)\} \geq \beta u(m) > \max\{v(m'|1), v(m'|0)\}$$

The second inequality is because $v(m'|s)$ is bounded above by $\max\{\beta u(m'), \beta u(m'+1) - 1\} < \beta u(m)$, as σ_D has to either follow m' for certain or mix with a neighboring action, depending on s.

Since $\max\{v(m|1), v(m|0)\} > \max\{v(m'|1), v(m'|0)\}$, we can conclude that $v(m|1) - v(m|0) = v(m'|1) - v(m'|0)$ implies $v(m|0) \neq v(m'|0)$. In other words, if two of the lines have the same slopes, they have different intercepts: They cannot be identical.

Observations 1 and 2 (i.e., the $v(m,p)$ is linear in p and non-identical for any pair m, m') allows the following conclusion: For all $m \in \tilde{M}$, the set $\{p \in [0,1] : v(m,p) \in max_{m'} v(m',p)\}$ is a closed and convex subset of $[0,1]$. Calling this interval $[\underline{p}_m, \bar{p}_m]$, m is the unique maximizer of $v(m,p)$ for $p \in (\underline{p}_m, \bar{p}_m)$.

Hence, for any given σ_D that has the above structure, the optimality of sending m holds only for a set of consecutive states ω. There is no state such that more than 2 messages are optimal. Furthermore, m is the unique maximizer of the expert utility in the states that are neither the highest or the lowest where m is optimal. This leads precisely to the structure described in Step 2.

Finally, note that the partition σ_E is *ordered* in the sense that $m > m'$ implies $\underline{\omega}_m \geq \bar{\omega}_{m'}$. To see this, note that for all $m \in \tilde{M}$, we must have $m \in \{\underline{\omega}_m, \ldots, \bar{\omega}_m\}$: Otherwise, message m can come from states that are exclusively larger or those that are exclusively smaller than $\omega = m$ due to the partition structure. Then, however, there exists $s \in \{0, 1\}$ such that the posterior mean $\tilde{\mu}(m, s) \notin [m - 0.5, m + 0.5]$ and therefore, D optimally responds with $\sigma_{m,s}(m) = 0$, contradicting the "never ignore for certain" property we established in Step 1.

Hence, the order of magnitude of message m among the on-path messages corresponds to the order of the state partition it is sent under.

Step 3: Candidate $\bar{\sigma}_E$ is the sender preferred partition: Finally, we show that among the remaining candidates, $\bar{\sigma}_E$ yields the highest ex-ante payoff for E for all η sufficiently small.

First, without loss, index the messages $m \in \tilde{M}$ on the equilibrium path by their magnitude. That is, let $m_1 < m_2 < ... < m_{|\tilde{M}|}$. We know, from the definition of ω_m in Step 2, the following observation on the incentive constraints for any $k, l \in \{1, ..., |\tilde{M}|\}$ with $k \neq l$:

$$p_\omega v(m_k|1) + (1 - p_\omega)v(m_k|0) \geq p_\omega v(m_l|1) + (1 - p_\omega)v(m_l|0) \qquad \text{for all } \omega \in \{\underline{\omega}_{m_k}, \ldots, \bar{\omega}_{m_k}\}$$

and the direction of the inequality is reversed for all $\omega \in \{\underline{\omega}_{m_l}, \ldots, \bar{\omega}_{m_l}\}$. For the strategy profile to be an equilibrium for arbitrarily small η, both of these inequalities have to maintain in the limit. This is only possible if the following equality holds for all $k \neq l$:

$$\begin{aligned} &\lim_{\eta\to 0}\big(p_\omega v(m_k|1) + (1 - p_\omega)v(m_k|0)\big) = \frac{1}{2}\lim_{\eta\to 0}\big(v(m_k|1) + v(m_k|0)\big) \\ &= \lim_{\eta\to 0}\big(p_\omega v(m_l|1) + (1 - p_\omega)v(m_l|0)\big) = \frac{1}{2}\lim_{\eta\to 0}\big(v(m_l|1) + v(m_l|0)\big) \qquad \text{(C)} \\ &\iff \lim_{\eta\to 0}\big(v(m_k|1) + v(m_k|0)\big) = \lim_{\eta\to 0}\big(v(m_l|1) + v(m_l|0)\big) \end{aligned}$$

This is a necessary condition because if one of the limits of these two (continuous) functions is strictly greater in the limit (e.g., greater for k), then the incentive constraint will be violated for the support states of the other (e.g., violated for $\{\underline{\omega}_{m_l}, \ldots, \bar{\omega}_{m_l}\}$).

With this in mind, we can write the limit of the ex-ante utility of E from profile σ with partition σ_E as follows. Taking $\bar{\omega}_{m_0} := 0$, we have:

$$
\begin{aligned}
\mathbb{E}_\sigma U_E &= \frac{1}{N} \sum_{m=m_1}^{m_{|\tilde{M}|}} \Bigg(\sum_{\omega=\bar{\omega}_{m-1}+1}^{\bar{\omega}_m} p_\omega v(m|1) + (1-p_\omega) v(m|0) \Bigg) \\
&\xrightarrow{\eta \to 0} \frac{1}{2N} \sum_{m=m_1}^{m_{|\tilde{M}|}} \Bigg(\sum_{\omega=\bar{\omega}_{m-1}+1}^{\bar{\omega}_m} \lim_{\eta \to 0} (v(m|1) + v(m|0)) \Bigg) \\
&= \frac{1}{2} \lim_{\eta \to 0} \Bigg(v(m_1|1) + v(m_1|0) \Bigg)
\end{aligned}
$$

where the equality is obtained by setting $k = 1$ in equation (C). Next, note that for β sufficiently small, we have $v(m_1|s) \leq \beta u(m_1)$ for all s.[26] That is, the limiting payoff from any candidate partition is bounded above by the payoff from its lowest message being followed for certain.

Finally note that if a non-pooling partition is incentive compatible, then in the limit, the highest action that it can induce from its lowest message m_1 with positive probability is $\frac{N-1}{2}$. To see this, note that no candidate equilibrium can have a message $m \in \tilde{M}$ such that $\sigma_\omega(m) = 0$ for all $\omega < N$. That is, there cannot be any message on path that is *only* sent in the highest state. This message would perfectly reveal state N, leading to optimal response of $\sigma_{m,s}(m) = 1$ for all s if $m = N$, and $\sigma_{m,s}(m) = 0$ for all s otherwise. In other words, it is either followed for certain with any private signal (if it is equal N), and ignored for certain (otherwise). Either way, this contradicts the necessary structure of "mix with a neighbor after one private signal, follow for certain after the other" we derived in Step 2.

This means the highest message must be sent with positive probability in state $N-1$. Therefore the interim mean induced by the lowest message m_1 is strictly below $\frac{N}{2}$.[27] Posterior mean $\frac{N}{2}$ is the indifference point for D between actions $\frac{N-1}{2}$ and $\frac{N+1}{2}$. So as $\eta \to 0$, any s will produce a posterior mean strictly below $\frac{N}{2}$, yielding an action $a \leq \frac{N-1}{2}$.

---

[26] We established in step two that D must respond to message m_1 with $a = m_1$, a neighboring action, or a mixture of these. Therefore, the specific "sufficiently small β" is given by $\beta < \frac{1}{u(m_1+1)-u(m_1)}$, which means E is better off when D follows message m for certain than if she takes action $m+1$.

[27] Recall that the "interim mean" is the Bayesian belief of D after observing m and before observing s. $\frac{N}{2}$ is the interim mean if the expert sends the highest message in $\omega = N$ and the lowest message for certain in all other states

Since m_1 cannot induce an action greater than $\frac{N-1}{2}$, the limit value of $\frac{v(m_1|1)+v(m_1|0)}{2}$ and therefore limit of $\mathbb{E}_\sigma U_E$ is bounded above by $\beta u(\frac{N-1}{2})$. Then, if there is a unique σ_E that supports an equilibrium where the ex-ante expert payoff converges to $\beta u(\frac{N-1}{2})$, it is the unique sender-preferred equilibrium strategy for small η.

Payoff from the candidate: Now we return to our candidate equilibrium $\bar{\sigma}$: Recall that $\bar{\sigma}_{\frac{N-1}{2},s}(\frac{N-1}{2}) = 1$ for all $s \in \{0, 1\}$, and $\frac{N-1}{2}$ is the lowest on-path message. Therefore $v(m_1|s) = v(\frac{N-1}{2}|s) = \beta u(\frac{N-1}{2})$ (and is independent of η) in this equilibrium. This means:

$$\lim_{\eta \to 0} \mathbb{E}_{\bar{\sigma}} U_E = \frac{v(\frac{N-1}{2}|1) + v(\frac{N-1}{2}|0)}{2} = \beta u\left(\frac{N-1}{2}\right)$$

so the expert payoff indeed converges to this highest feasible upper bound in equilibrium $\bar{\sigma}$.

Uniqueness: Now we show that $\bar{\sigma}_E$ is the unique expert strategy that sustains an equilibrium which achieves this payoff in the limit. First note that that if lowest on-path message m_1 achieves payoff $\beta u(\frac{N-1}{2})$, then it must itself be $m_1 = \frac{N-1}{2}$. This is because We already established that the highest action m_1 can induce is $a = \frac{N-1}{2}$, which means the payoff from any message $m \neq \frac{N-1}{2}$ is necessarily bounded above by $\beta u(\frac{N-1}{2}) - 1$ (i.e., it yields the "mismatch penalty").

Second, for $m_1 = \frac{N-1}{2}$ to be met with response $a = m$ and yield payoff $\beta u(\frac{N-1}{2})$, the posterior mean it induces must satisfy $\tilde{\mu}(m_1, s) \in (\frac{N-2}{2}, \frac{N}{2})$ for all s. Due to the partition structure of candidate profiles, this means for small η, m_1 must be sent for certain in states $\omega \in \{1, \ldots, N-2\}$.

Two summarize the two observations, any candidate for a sender preferred non-pooling equilibrium must yield $\sigma_\omega(\frac{N-1}{2}) = 1$ for all $\omega \in \{1, \ldots, N-2\}$. The remaining conditions shall determine the actions in states $N-1$ and N.

First, recall from Step 2 that partitions can overlap at most in one state. Therefore, there cannot be m, m' such that $\min\{\sigma_N(m), \sigma_N(m'), \sigma_{N-1}(m), \sigma_{N-1}(m')\} > 0$. That is, E cannot mix between a given pair of messages in both states $N-1$ and N.

Second, no message $m \neq m_1$ can perfectly reveal either state $\omega \in \{N-1, N\}$, because then we would have $m > m_1$ with $\sigma_{\omega,s}(m) = 1$ for all s, a property ruled out for candidates in Step 2. So if a message $m \neq m'$ is sent under one of these states, it should be sent under both.

Hence there can be only a single message $m' > m_1$ that is sent with positive probability in these states. The only way to avoid mixing m' with $\frac{N-1}{2}$ in multiple states while preserving the partition structure is if $\sigma_N(m') = 1$. That is, we are looking for a profile with $\sigma_\omega(\frac{N-1}{2}) = 1$ for all $\omega \in \{1, \ldots, N-2\}$, and there exists a unique $m' > \frac{N-1}{2}$ such that $\sigma_{N-1}(m') > 0$, $\sigma_N(m') = 1$ and $\sigma_\omega(m) = 0$ if $m \notin \{m', \frac{N-1}{2}\}$ for $\omega \in \{N-1, N\}$.[28]

We established in Step 2 that $m \in \{\underline{\omega}_m, \ldots, \bar{\omega}_m\}$. Therefore, we must have $m' \in \{N-1, N\}$. However, $\sigma_N(m') = 1$ together with $\sigma_\omega(m') = 0$ for all $\omega < N-1$ means that $\tilde{\mu}(m', 1) > N - \frac{1}{2}$ and therefore $\sigma_{m',1}(N) = 1$. So to maintain the "never ignore for certain" property established in Step 1, we must have $m' = N$. So the sender-preferred strategy of the expert must send message $\frac{N-1}{2}$ for certain under $\omega \in \{1, \ldots, N-2\}$, message N for certain under $\omega = N$, and message N with positive probability under $\omega = N-1$.

Finally, we need to check which mixes between the two possible messages in state $N-1$ (i.e., $\sigma_{N-1}(N)$) sustain an equilibrium. Recall that $q_E := \frac{1-p_N}{1-p_{N-1}}$ as derived in candidate equilibrium $\bar{\sigma}$ solves $\tilde{\mu}(N, 0) = N - \frac{1}{2}$. So if $\sigma_{N-1}(N) < q_E$, then we have $\tilde{\mu}(N, 0) < N - \frac{1}{2}$ and therefore $\sigma_{N,0}(N) = 0$, violating the "never ignore for certain" condition. If $\sigma_{N-1}(N) > q_E$, then we have $\tilde{\mu}(N, 0) > N - \frac{1}{2}$ yielding $\sigma_{N,0}(N) = \sigma_{N,1}(N) = 1$, so message N is followed for certain without being the lowest message in $\tilde{M}$, which is ruled out in Step 2. Hence, the only possible mix is $\sigma_{N-1}(N) = q_E$. This means the only possible expert strategy is $\bar{\sigma}_E$ as defined in candidate equilibrium $\bar{\sigma}$.

Optimal Commitment Strategy

Now we turn to the commitment problem and derive the structure of the optimal message strategy for convex u, before comparing it to the sender-preferred influence motivated equilibrium.

With uninformed D, the commitment problem is fairly simple and we can use the standard concavification approach. First note that since $U_D = -(a - \omega)^2$, the best response of D is to simply choose the action closest to posterior mean $\tilde{\mu}(m)$, and if $\tilde{\mu}(m) = \omega - 0.5$ for some ω, choose $a = \omega$ (i.e., expert preferred action). Call this $a^*(\tilde{\mu})$.

[28]Note that the case $m' = \frac{N-1}{2}$, corresponds to a "pooling strategy" where $\frac{N-1}{2}$ is sent for certain in all states. Therefore, we can rule it out for the statement of the Theorem.

To formally define the problem, denote by $\mu(\omega)$ the posterior belief that the state is ω. Since we are interested in the optimal *binary* signal, the problem is to choose two feasible posterior means that maximize the expert value. The expert problem can be written in terms of induced posterior means as follows:

$$\begin{aligned} \max_{\tilde{\mu}_L,\tilde{\mu}_H} \quad & (1-p)\,u(a^*(\tilde{\mu}_L)) + p\,u(a^*(\tilde{\mu}_H)) \\ \text{s.t.} \quad & \exists \mu_L(\cdot), \mu_H(\cdot) \in \Delta(\{1,...,N\}) \text{ with} \\ & \tilde{\mu}_L = \sum_{\omega=1}^{N} \omega\, \mu_L(\omega), \quad \tilde{\mu}_H = \sum_{\omega=1}^{N} \omega\, \mu_H(\omega), \\ & (1-p)\,\mu_L(\omega) + p\,\mu_H(\omega) = \frac{1}{N}, \quad \forall \omega = 1, \ldots, N. \end{aligned}$$

where the constraints state that $\tilde{\mu}_H$ and $\tilde{\mu}_L$ are means induced by two Bayes plausible posterior beliefs, where "high belief" μ_H is induced with probability p. In other words, these constraints describe the pairs of posterior means that are *feasible under some binary message strategy*. The posterior means themselves have to average out to the prior mean, so we can write:

$$(1-p)\tilde{\mu}_L + p\tilde{\mu}_H = \frac{N+1}{2}$$

We can solve this "mean plausibility" constraint for p and rewrite the objective as

$$\max_{\tilde{\mu}_L,\tilde{\mu}_H} \quad \frac{\tilde{\mu}_H - \frac{N+1}{2}}{\tilde{\mu}_H - \tilde{\mu}_L} u(a^*(\tilde{\mu}_L)) + \frac{\frac{N+1}{2} - \tilde{\mu}_L}{\tilde{\mu}_H - \tilde{\mu}_L} u(a^*(\tilde{\mu}_H))$$

subject to the feasibility constraints. Note now that as $\min_{a\in\{2,...,N-1\}} \Delta_u^2(a)$ goes to infinity, $u(a)$ goes to 0 for all $a < N$, and $u(N)$ goes to 1.[29] So we have: if $\tilde{\mu}_H < N - 0.5$, then the objective function converges to zero with any feasible $\tilde{\mu}_L$ as $\min_{a\in\{2,...,N-1\}} \Delta_u^2(a)$ goes to infinity. Clearly, for sufficiently convex u, such a strategy is suboptimal.

If we set $\tilde{\mu}_H \geq N - 0.5$, then the objective function converges to $p = \frac{\frac{N+1}{2} - \tilde{\mu}_L}{\tilde{\mu}_H - \tilde{\mu}_L}$ as $\min_{a\in\{2,...,N-1\}} \Delta_u^2(a)$ goes to infinity. Therefore, for sufficiently convex u, the objective of E is to maximize the probability p that a high posterior mean with $\tilde{\mu}_H \geq N - 0.5$ is induced. This probability is strictly decreasing in both terms, so it is maximized when (i) For the high posterior mean $\tilde{\mu}_H = N - 0.5$ constraint is binding. (ii) Low posterior mean

[29] Recall that we have fixed $u(1) = 0$ and $u(N) = 1$

$\tilde{\mu}_L$ is the minimum value that can be feasibly paired with $N - 0.5$ under a binary signal. This value is given by $\tilde{\mu}_L = \frac{N-1}{2}$, obtained through a message strategy that sends m for certain under states $\omega \in \{1, ..., N-2\}$, inducing beliefs $\mu_L(\omega) = \frac{1}{N-2}$ for each. The complementary message $m' \neq m$ is sent for certain under states $N - 1$ and N, inducing beliefs $\mu_H(N) = \mu_H(N-1) = 0.5$.

Recall that our sender preferred influence motivated equilibrium strategy σ_E converges with $\eta \to 0$ to $\bar{\sigma}_\omega(\frac{N-1}{2}) = 1$ for all $\omega \leq N - 2$ and $\bar{\sigma}_{N-1}(N) = \bar{\sigma}_N(N) = 1$. Thus, setting $m = \frac{N-1}{2}$ and $m' = N$ yields the limiting equivalence.

□

References

Boleslavsky, R., & Shadmehr, M. (2024). *Signaling With Commitment* (No. 2305.00777).

Chakraborty, A., & Harbaugh, R. (2010). Persuasion by cheap talk. *American Economic Review, 100*(5), 2361-2382.

Chen, Y. (2012). Value of public information in sender–receiver games. *Economics Letters, 114*(3), 343-345.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431-1451.

Dziuda, W., & Salas, C. (2018). Communication with detectable deceit. *Available at SSRN 3234695.*

Gallice, A., & Grillo, E. (2025). Shifting social norms through endorsements. *Journal of Economic Behavior & Organization, 231*, 106903.

Ishida, J., & Shimizu, T. (2016). Cheap talk with an informed receiver. *Economic Theory Bulletin, 4*, 61-72.

Kamenica, E., & Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review, 101*(6), 2590-2615.

Koessler, F., Laclau, M., & Tomala, T. (2024). *A belief-based approach to signaling.* HAL.

Kolotilin, A., Mylovanov, T., Zapechelnyuk, A., & Li, M. (2017). Persuasion of a privately informed receiver. *Econometrica, 85*(6), 1949-1964.

Kolotilin, A. (2018). Optimal information disclosure: A linear programming approach. *Theoretical Economics, 13*(2), 607-635.

Lai, E. K. (2014). Expert advice for amateurs. *Journal of Economic Behavior & Organization, 103*, 1-16.

Lipnowski, E., & Ravid, D. (2020). Cheap talk with transparent motives. *Econometrica, 88*(4), 1631-1660.

Mathevet, L., Pearce, D., & Stacchetti, E. (2022). *Reputation for a Degree of Honesty*. Working paper.

Olszewski, W. (2004). Informal communication. *Journal of Economic Theory, 117*(2), 180-200.

Ottaviani, M., & Sørensen, P. N. (2006). Reputational cheap talk. *The Rand journal of economics, 37*(1), 155-175.