# Cheap Talk with Costly Verification

May 14, 2025

**Abstract**

I study a model of cheap talk between a sender and receiver. The sender, who is informed about the state of the world, communicates with a receiver, after which the receiver decides whether to pay a cost, the size of which is private information, to learn whether the sender's message is true. I show that any influential equilibrium is characterized by a threshold. If the state is above the threshold, the sender tells the truth, and if the state is below the threshold, the sender lies by pretending to be a truth-teller. In response, the receiver verifies all messages above the threshold with positive probability. I then show that receiver's ability to verify the sender's message always has an informational effect, but it sometimes also has a deterrence effect. Moreover, the deterrence effect can increase the expected informational effect of verification.

# 1 Introduction

In the early months of the Covid-19 pandemic, Vice President Mike Pence held a press conference as member of President Trump's Coronavirus Task Force. At the press conference, he said:

> "...[w]e want the American people to understand it's almost inarguable that more testing is generating more cases. To one extent or another the volume of new cases coming in is a reflection of a great success in expanding testing across the country."[1]

That same day, Linda Qui, a fact-checking reporter at the *The New York Times*, published an article in which she deemed Vice President Pence's statement "false." Had interested readers sought our her article, they would have read her analysis:

> "Ramped up testing alone does not account for the uptick in cases. Rather, the virus's spread is generating more cases..."[2]

She supported her verdict by quoting Dr. Robert R. Redfield, the director of the Centers for Disease Control and Prevention, who had recently said

> "Several communities are seeing increased cases driven by multiple factors, including increased testing, outbreaks and evidence of community transmission."[3]

In this paper, I study communication between a sender and receiver when the receiver can pay a cost to learn about the veracity of the sender's message. To do so, I focus on two key aspects illustrated by the previous anecdote. First, while information confirming whether the sender's statement is true may be available, the receiver must endogenously choose to access this information to learn from it. While there is a robust network of fact-checkers who respond to claims made by politicians—journalists at places like *The New York Times* or

[1] Pence (2020)
[2] Qui (2020)
[3] Qui (2020)

the *Washington Post*, employees of non-profits like *FactCheck.org*, etc.—a voter must choose to search for this information to learn whether a politician's statement is true. Second, the information available about the sender's message may only reveal whether the message is true, not the actual state of the world. Linda Qui's article deemed Pence's statement false, but it did not reveal the state of the world: the Trump administration's responsibility for the spread of Covid-19.

In particular, I study a model of cheap talk with endogenous verification. At the start of the game, the sender (she) learns the state of the world, and the receiver (he) learns how costly it will be to verify the sender's message. The sender communicates with the receiver by choosing a message, after which the receiver chooses whether to pay his cost to verify the message. If he does, he learns whether it is true but does not learn the state of the world. If he does not, he learns nothing about its veracity. Then, the receiver chooses an action that is payoff relevant to both players.

In my first primary result, I show that any influential equilibrium of the model has a straightforward structure characterized by a threshold, the sender's expected utility in equilibrium when she lies. When the state is above this threshold, the sender tells the truth, and when the state is below this threshold, she lies by pretending to be a high type. Aside from one knife-edge case, any message the sender sends truthfully, she also sends as a lie, and the receiver verifies all messages in the support of the sender's strategy with positive probability.

This equilibrium structure reveals a dependence between endogenous verification and lying. If the receiver believes a particular message in the support of the sender's strategy is only sent truthfully, he has no incentive to verify it. But, as described above, in any influential equilibrium, the messages in the support of the sender's strategy are all above a threshold, the sender's expected utility in equilibrium when she lies. So if there is a message in the support of the sender's strategy that is only sent truthfully, a lying sender has a profitable deviation to the message that is never verified. This dependence illustrates a key difference between endogenous verification, studied in this model, and exogenous verification, studied elsewhere (e.g., ?), and illustrates how endogenous verification relies on the expectation of lying.

I then characterize all influential equilibria when there are three equally likely states and the cost of verification is uniform. This characterization illustrates the two effects of verification. The first is the direct *informational* effect, which is the reduction in the variance of the receiver's belief he obtains through verification. The second is the indirect deterrence effect of verification: verification can deter the intermediate type from lying due to an endogenous cost of lying that emerges in equilibrium. Together, these effects mean that in any influential equilibrium, the receiver's expected utility is higher than what he obtains without this verification technology.

One might conjecture that the informational and deterrence effects act like substitutes: the greater the deterrence effect, the less the informational effect. I conclude by analyzing a region of the parameter space where two equilibria exist, one where verification only has an informational effect and one where verification has an informational and deterrence effect. In my second primary result, I show previous intuition is not always correct. In particular, in some cases, the expected informational effect of verification is greater when verification has a deterrence effect than when it does not.

## 2 Related Literature

This paper contributes to the literature on communication with detectable lying, particularly in settings where communication takes the form of cheap talk. Some existing work studies cheap talk communication when the sender's message is verified with an exogenous probability (e.g., Dziuda and Salas, 2018; Balbuzanov, 2019; Holm, 2010).[4] In other work, as in this paper, the receiver endogenously chooses whether to verify the sender's message

[4] Less related to this model, others study exogenously detected deceit in other communication frameworks like Bayesian persuasion (e.g., Ederer and Min, 2022; Venkatesh et al., 2025)

(e.g., Sadakane and Tam, 2023; Levkun, 2022; Ball and Gao, 2025).

In Sadakane and Tam (2023), the paper closest to mine, a receiver chooses whether to incur a publicly known cost to inspect a message sent by a privately informed sender. In my model, less punishing off the equilibrium path beliefs mean that in any IE, except in one knife-edge case, all messages in the support of the sender's strategy are sent truthfully and as lies and are verified with positive probability. This is not the case in Sadakane and Tam (2023), where if the state is uniformly distributed, some messages are not verified. In both models, the receiver's ability to verify the sender's message has two effects: an informational effect and a deterrence effect. However, in Sadakane and Tam (2023), the informational effect is completely offset by the cost of verification, which leaves the authors pessimistic about the informational benefit of fact-checking. In contrast, when the receiver has private information about his cost, the informational effect exceeds the ex-ante expected cost of verification, suggesting a rosier picture. Moreover, because the cost in my model does not offset the informational effect, I show that, in some cases, the deterrence effect can increase the expected informational effect.

My model shares a similar structure with Dziuda and Salas (2018) but is distinct because the receiver endogenously chooses whether to verify the sender's message. This endogenous choice produces a distinct equilibrium structure. In Dziuda and Salas (2018), some intermediate messages are only sent truthfully, whereas, in my model, all messages in the support of the sender's strategy are sent truthfully and as lies.[5] When verification is exogenous, messages that are only sent truthfully on the equilibrium path are still verified; this prevents the sender from deviating to lying with one of these messages. When verification is endogenous, messages only sent truthfully on the equilibrium path are never verified, giving the sender a profitable deviation. Hence, my model illustrates how the effectiveness of fact-checking relies on the expectation of politicians lying.

This paper is also related to models of cheap talk with an exogenous cost of lying. This

[5]Except in a knife-edge case discussed below.

cost might arise due to a psychological cost associated with lying (e.g., Kartik et al., 2007) or the existence of receivers who take the sender's message at face value (e.g., Kartik et al., 2007).[6] In both cases, deception emerges in equilibrium, but the cost of lying constrains the sender to the extent that information can be conveyed despite disagreement between the sender and receiver(s). In my model, the receiver does not pay an exogenous cost to lie. Yet, I show that in equilibrium, an endogenous cost of lying emerges: because verification does not reveal the state, a sender caught lying will be pooled with the other types who lie. Because low types lie, the cost of being pooled with other liars can deter higher types from lying.

Finally, this paper is related to the literature on fact-checking. This literature is primarily empirical (e.g., Weeks and Garrett, 2014; Weeks, 2015; Nyhan and Reifler, 2010), but includes some theoretical work (e.g., Levkun, 2022). Of particular relevance to my paper are Nyhan and Reifler (2015) and Lim (2018), who document empirical evidence of the direct benefit of fact-checking on politicians' behavior, and Gottfried et al. (2013) and Pingree et al. (2014), who document empirical evidence of the direct benefit of fact-checking on voter's information.

# 3 Model

There is a sender ($S$, "she") and a receiver ($R$, "he"). At the start of the game, the sender privately learns the state of the world, her "type", $\theta \in \Theta$, where $\Theta$ is a finite set of real numbers, each $\theta \in \Theta$ occurs with probability $h(\theta) > 0$, $N = |\Theta|$, and $\theta \in \Theta$ are indexed such that $\theta_1 < ... < \theta_N$. The sender communicates with the receiver by choosing a message $m$ from the message space $M = \theta$. If $m = \theta$, the message is "true" (the sender tells the truth), and if $m \neq \theta$, the message is "false" (the sender lies).[7]

After observing the sender's message, the receiver decides whether to verify it ($c = 1$) or

[6] To a lesser degree, my model is related to a broader literature of strategic communication where lying is costly (e.g., Nguyen and Tan, 2021; Guo and Shmaya, 2019).

[7] Since the sender is restricted to choosing a single message rather than allowed to choose a subset of the message space (e.g., $m = \{\theta_j, \theta_k\}$), these definitions are equivalent to the definitions in Sobel (2020).

not ($c = 0$). Verification is costly; if the receiver verifies, he pays the cost $\beta\kappa$ where $\kappa \in [0, 1]$ is private information he learns at the start of the game and $\beta > 0$ is a parameter observed by both players. I assume $\kappa$ has a prior distribution $g$ and that $g(\kappa) > 0$ for all $\kappa \in [0, 1]$. If the receiver verifies the sender's message, he learns whether it is true ($v = t$) or false ($v = f$) but does not learn the state. If he does not verify, he learns nothing ($v = \emptyset$). Finally, the receiver selects an action $a \in [0, 1]$.

**Preferences.** The receiver has a utility function $u_R(\kappa, \theta, a) = -(\theta - a)^2 - c \cdot \beta\kappa$, and the sender has a utility function $u_S(a) = a$. Note, the receiver's payoff depends on the state—in particular, he wants choose the action that matches the state—while the sender wants the receiver to choose $a = 1$.

**Equilibrium.** A (mixed) strategy for the sender is a probability function $\sigma(\cdot|\theta) : \Theta \to \Delta M$.[8] A message $m$ is "on the equilibrium path" or in "the support of" the sender's strategy if $\sigma(m|\theta) > 0$ for at least one $\theta$, and is "off the equilibrium path" otherwise. A strategy for the receiver is a tuple $(c, a)$ where $c : K \times M \to \{0, 1\}$ and $a : M \times \{t, f, \emptyset\} \to [0, 1]$.[9] The game is solved for perfect Bayesian equilibrium (henceforth, an "equilibrium"), where a PBE is a triple $(\sigma, c, a)$ and a belief assessment such that:

1. If $m \in M$ is in the support of $\sigma(m|\theta)$, $m$ maximizes the sender's expected utility taking $(c, a)$ as given.

2. If $(c, a)$ is chosen, then $c$ maximizes the receiver's expected utility given $\mu_1 = \mathbb{E}[\theta|m]$ and $a$ maximizes the receiver's expected utility given $\mu_2 = \mathbb{E}[\theta|(m, v)]$.

3. Beliefs $\mu_1$ and $\mu_2$ are generated using Bayes' rule on the equilibrium path.

Throughout the paper, I focus on solving for influential equilibria (henceforth, IEs).

[8] $\Delta(X)$ denotes the space of lotteries over $X$.

[9] I restrict attention to equilibria where the receiver plays a pure strategy. This is without loss of generality as the receiver has a unique best response to any posterior belief about the sender's type due to the strict concavity of his utility function.

**Definition 1.** *An equilibrium is influential if $a^*(m, v)$ is not constant on the equilibrium path (Sobel, 2013).*

**Assumptions.** As in many cheap talk games, the set of equilibria is large. In light of this, I make the following assumptions to refine the set of equilibria.

**Assumption 1.** *When lying, the sender's strategy is independent of $\theta$.*

In equilibrium, when the sender lies, she must be indifferent between any lies within the support of her strategy. Otherwise, she has a profitable deviation. Therefore, an equilibrium where the sender conditions her lying strategy on $\theta$ requires different types of senders to use different strategies to lie even though each strategy yields the same expected utility. To simplify the analysis of the game, I assume the sender's strategy is independent of $\theta$ when she lies.

**Assumption 2.** *If there is a message $m$ that is not in the support of the sender's strategy, then upon seeing $m$, the receiver believes $\theta = m$ and chooses $a = m$.*

This assumption can be justified by assuming that if a message is off the equilibrium path, there is an infinitesimally small probability the truth is accidentally revealed by the sender through a slip of the tongue (e.g., Hart et al., 2017).

**Discussion of the Model.** This paper studies a model of communication between a sender and receiver where the receiver has access to verification technology that reveals whether the sender's message is true, and accessing it requires paying a privately known cost. This model applies naturally to political communication, where a politician makes a statement, a journalist fact-checks the statement, and a voter or other interested party who observed the initial statement chooses whether to pay a cost to access the journalist's fact-checking assessment. It also applies to settings outside of politics like job interviews, where an interviewee makes statements about her qualifications (e.g., "At my last job, I was responsible for this task...”), and the interviewer can verify whether this information is true by conducting a background check or speaking to the interviewee’s former employer.

A key feature of this model is that when the sender makes her statement, she does not know the cost the receiver will need to pay to verify it. This uncertainty may arise for many reasons. In the political communication example, the politician may be uncertain which news sources or how many will fact-check her statement. If more sources fact-check the statement, it is cheaper for the voter to verify. On the other hand, if the only sources who fact-check have a paywall, it may be very costly for the voter to verify. In the job interview example, the interviewee may be unsure how badly the interviewer needs to fill the position, which affects how costly it is to conduct a background check.

While the sender does not know the cost the receiver will pay to verify her statement, she knows $\beta$, which is the upper bound of possible costs. In the political communication example, this upper bound represents features of the political environment that affect the distribution of possible costs. For instance, a politician running in a local race may be much less likely to face fact-checking, and as a result, $\beta$ might be very large. On the other hand, a national politician like the president might be extremely likely to be fact-checked, and as a result, $\beta$ might be very small.

Another key feature of the model is that while verification reveals whether the sender’s message is true, it does not reveal the state. Many political fact-checking sources use a shorthand to grade the veracity of a statement. For example, the *Washington Post* uses Pinocchios to convey whether a message is true or false, and *The New York Times* determines whether a message is *false*, *true*, *misleading*, etc. While these grading systems convey information about the politician’s statement, they do not necessarily reveal the state of the world. This is illustrated by the anecdote at the beginning of the paper where Linda Qui’s coverage deemed Mike Pence’s statement false but could not determine their responsibility of the Trump administration for the Covid-19 pandemic.

# 4 Analysis

## 4.1 Equilibrium Structure

I begin by considering the receiver's action as a function of the message he observes and any additional information learned from verification. In the final move of the game, he solves:

$$\max_{a\in[0,1]} -\mathbb{E}[(\theta - a)^2|(m,v)]. \tag{1}$$

Expression (1) is uniquely maximized when $a^*(m,v) = \mathbb{E}[\theta|(m,v)]$.

Before choosing his action, the receiver decides whether to verify $m$. The optimality of $a^*(m,v)$ and the fact that the receiver has quadratic loss utility means he will verify $m$ if:

$$\begin{aligned} &Var(m|c=0) - \phi(m)Var(m|c=1, v=t) - (1-\phi(m))Var(m|c=1, v=f) \\ =&\underbrace{Var(m|c=0) - (1-\phi(m))Var(m|c=1, v=f)}_{\textbf{Informational Effect of Verification: } \Lambda(m)} > \beta\kappa, \end{aligned} \tag{2}$$

where $Var(m|\cdot)$ is the variance of the receiver's belief, $\phi(m)$ is the voter's conjecture about the probability $m$ was sent truthfully, and the second line follows from the fact that $Var(m|c = 1, v = t) = 0$. The left-hand side of (2) represents the informational effect of verification, which I denote $\Lambda(m)$. Intuitively, for the sender to verify $m$, the informational effect must exceed the cost. From (2), it is clear the receiver will not verify a message he believes is only sent truthfully, nor will he verify a message he believes is only sent as a lie since $\Lambda(m) = 0$ is zero in both cases.

In any IE, some types of sender must lie. Suppose not, then the receiver believes the sender's message and does not verify. However, if this is the case, a low-type sender has a profitable deviation to sending a higher message. Hence, in any IE, at least one type of sender lies, and she receives an expected payoff $u^{lie}$. Assumption 1 implies that $u^{lie}$ is independent of $\theta$.

In the following proposition, I provide the structure of any IE of this game.

**Proposition 1.** *In any IE, the sender tells the truth if $\theta > u^{lie}$ and lies if $\theta < u^{lie}$; when the sender lies, she lies by randomizing over each message $m > u^{lie}$ with positive probability; and the receiver verifies all messages $m > u^{lie}$ with positive probability.*

Proposition 1 states that any IE has a simple structure: the sender tells the truth when she has a high type and lies when she has a low type by randomizing over messages that are sent truthfully, and the receiver verifies all messages $m > u_{lie}$ with positive probability.[10] Moreover, except in the knife-edge case where $\theta = u^{lie}$, any message that is sent truthfully is sent as a lie and verified with positive probability.

The intuition for this structure is as follows. As discussed above, in any IE, at least one type of sender must lie. Moreover, all messages $m \neq u^{lie}$ that are sent truthfully are also sent as lies.[11] Suppose not. Then, there is a message $m$ that is only sent truthfully. The receiver will not verify $m$ and will choose $a(m, \emptyset) = m$. Either $m > u^{lie}$ or $m < u^{lie}$. The sender can deviate from lying to sending $m$ in the former. In the latter, the sender can deviate from telling the truth with $m$ to lying.

Consider the sender with type $\theta > u^{lie}$. Suppose she lies. As a result, she receives an expected payoff of $u^{lie}$. Since the type-$\theta$ sender lies, $m = \theta$ must be off the equilibrium path. This is because Assumption 1 implies all senders who lie use the same strategy, meaning $m = \theta$ cannot be sent as a lie. Since $m = \theta$ is off the equilibrium path, Assumption 2 implies the sender can deviate from lying to reporting $m = \theta$, which will be believed. Thus, senders of type $\theta > u^{lie}$ tell the truth.

Consider the sender with type $\theta < u^{lie}$, and suppose she tells the truth. When the receiver verifies $m = \theta$, he chooses $a(m, t) = \theta < u^{lie}$. As discussed above, it must be that $m = \theta$ is also sent as a lie. This implies the action the receiver takes if he does not verify

[10] Proposition 1 provides a characterization, not existence. Without more structure on $\Theta$, it is difficult to derive a necessary and sufficient condition for existence. However, in the Supplementary Appendix, I provide sufficient conditions for some IEs.

[11] In the Appendix, I address the knife-edge case where $\theta = u^{lie}$.

$m = \theta$ is also smaller than $u^{lie}$ since the sender tells the truth if $\theta > u^{lie}$. Hence, the sender with type $\theta < u^{lie}$ who tells the truth has a profitable deviation.

**Equilibrium Structure** The structure of any IE in this model is similar to the structure of equilibria in other models where the sender's lies may be detected through exogenous verification (e.g., Dziuda and Salas, 2018) or through endogenous verification (e.g., Sadakane and Tam, 2023). Across these models, high types tell the truth, and low types lie by pretending to be high types. Yet, the equilibria in Dziuda and Salas (2018) and Sadakane and Tam (2023) differ in one critical way: not all messages that are sent truthfully are sent as lies.[12]

Relative to Dziuda and Salas (2018), the difference emerges because of the difference between exogenous and endogenous verification. If the receiver believes a message is only sent truthfully, he will not pay the cost to verify it when verification is endogenous. But if he never verifies a message that is in in the support of the sender's strategy, the sender can deviate to that message without fear of getting caught. Hence, in equilibrium, the receiver must verify all messages in the support of the sender's strategy.[13] But the receiver cannot commit to verifying every message; he will only verify if doing so is sequentially rational. This requires that each message the sender sends truthfully, she also sends as a lie. In contrast, if a message is stochastically verified, it will be verified with positive probability even if it is only sent truthfully on the equilibrium path. This constrains the behavior of the sender. Thus, this comparison illustrates a sense in which successful fact-checking requires suspicion to be effective—otherwise, the receiver will not verify the sender's message.

Relative to Sadakane and Tam (2023), the difference emerges due to assumed off the equilibrium path beliefs. To see this, suppose that in my model, like Sadakane and Tam (2023), the sender can send a message $m = [\theta_j, \theta_k]$ with the meaning "the state is between" $\theta_j$ and $\theta_k$. If $m \in [\theta_j, \theta_k]$ this message is truthful and if $m \notin [\theta_j, \theta_k]$ this message is a lie.

[12]Note, in the knife-edge case where $u^{lie} = \theta$, the message $m = \theta$ may only be sent truthfully.
[13]As long as there is not an $m = u^{lie}$.

Furthermore, consider the following example:

**Example 1.** *$\Theta = \{0, \frac{1}{32}, \frac{3}{32}, 1\}$, $h(\theta) = \frac{1}{4}$ for all $\theta$, $\beta = 1$, $G = \mathcal{U}[0,1]$, the sender sends $m = 1$ if $\theta = 1$, $m = [\frac{1}{32}, \frac{3}{32}]$ if $\theta \in \{\frac{1}{32}, \frac{3}{32}\}$, and lies by reporting $m = 1$ if $\theta = 0$.*

When the receiver observes $m = 1$, he verifies when $\kappa < \frac{1}{4}$, and he never verifies $m = [\frac{1}{32}, \frac{3}{32}]$. Given such a strategy, the type-0 sender and the senders of types $\frac{1}{32}$ and $\frac{3}{32}$ are indifferent between sending $m = 1$ and $m = [\frac{1}{32}, \frac{3}{32}]$ because both messages yield an expected payoff of $\frac{1}{8}$. However, among other profitable deviations, the type-$\frac{3}{32}$ sender can deviate to $m = \frac{3}{32}$, which will be believed. In contrast, Sadakane and Tam (2023) assume the receiver believes a sender who deviates is the lowest type. This punishing belief can support this behavior in equilibrium.

## 4.2 Three-State Example

Proposition 1 described the structure of any IE. In this section, I focus on the case where $\Theta = \{0, \theta_2, 1\}$ with $\theta_2 \in (0, \frac{1}{2})$, $h(\theta) = \frac{1}{3}$ for all $\theta$, and $G = \mathcal{U}[0,1]$ (i.e., the distribution of the cost to verify is uniformly distributed on the unit interval). Note that in the traditional cheap talk model with these preliminaries, there is no influential equilibrium.[14]

Denote the probability the receiver verifies a particular message in the support of the sender's strategy as $\pi(m)$. The following proposition characterizes all IE.[15]

**Proposition 2.** *Define $\tilde{\beta}, \overline{\beta}$, and $\underline{\beta}$ as in (4), (6), and (8). An IE exists in all regions of the parameter space. In particular,*

*(a.) if $\beta > \tilde{\beta}$, an IE exists where the sender tells the truth when $\theta = 1$, lies when $\theta \in \{0, \theta_2\}$ by sending $m = 1$, and $\pi(1) \in (0, 1)$,*

[14] Suppose not, then there must be $m$ and $m'$ such that $a(m) \neq a(m')$. But then $a(m) > a(m')$ or $a(m) < a(m')$, which implies there is a profitable deviation.

[15] If $\theta_2 \in [\frac{1}{2}, 1)$, the IE described in Proposition 2($a$) does not exist and there is no upper bound on the value of $\beta$ such that the IE described in Proposition 2($b$) exists.

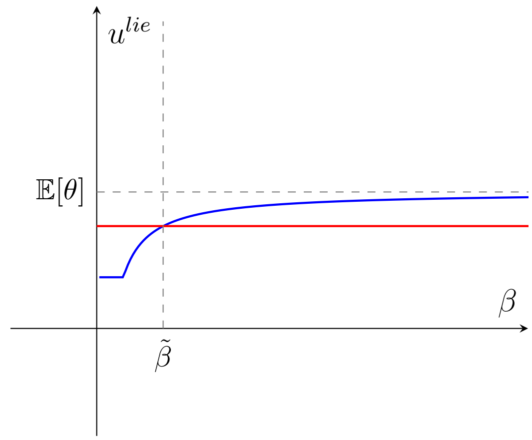


Figure 1: Expected utility from lying with $m = 1$ and from deviating to $m = \theta_2$, which is off the equilibrium path, for the sender of types $\theta \in \{0, \theta_2\}$. Assumes $\theta_2 = \frac{1}{3}$.

*(b.) if* $\beta \in (\underline{\beta}, \overline{\beta})$*, where* $\underline{\beta} < \tilde{\beta}$*, an IE exists where the sender tells the truth when* $\theta \in \{\theta_2, 1\}$*, lies when* $\theta = 0$ *by randomizing between* $m = 1$ *with probability* $\sigma^*$ *and* $m = \theta_2$ *with probability* $1 - \sigma^*$ *satisfying (5), and* $\pi(1), \pi(\theta_2) \in (0, 1)$*,*

*(c.) if* $\beta = \overline{\beta}$*, an IE exists where the sender tells the truth when* $\theta \in \{\theta_2, 1\}$*, lies when* $\theta = 0$ *by sending* $m = 1$*,* $\pi(\theta_2) = 0$ *and* $\pi(1) \in (0, 1)$*,*

*(d.) and if* $\beta < \underline{\beta}$*, a continuum of IEs exist where the sender tells the truth when* $\theta \in \{\theta_2, 1\}$*, lies when* $\theta = 0$ *by randomizing between* $m = 1$ *with probability* $\sigma^*$ *and* $m = \theta_2$ *with probability* $1 - \sigma^*$ *satisfying (9), and* $\pi(1) = \pi(\theta_2) = 1$*.*

Proposition 2 states that an IE exists in all regions of the parameter space. Intuitively, when $\beta$ is large, the receiver's verification is stochastic, but when $\beta$ is small, he verifies each message in the support of the sender's strategy with certainty.

Figure 1 provides an intuition for Proposition 2$(a)$ by depicting the type-0 and type-$\theta_2$ sender's expected utility from lying as a function of $\beta$. This expected utility is increasing in $\beta$ since the probability the receiver verifies is decreasing in $\beta$, and a lower probability of verification means a higher probability of successfully pooling with the type-1 sender. For this type of IE to exist, this expected utility must be weakly larger than $\theta_2$, which is the payoff that can be obtained by deviating to $m = \theta_2$, which is off the equilibrium path. Hence, when $\beta$ is sufficiently large, denoted by $\tilde{\beta}$ in the figure, this type of IE exists. Additionally, as $\beta$ goes to infinity, the probability of verification goes to zero, and $u^{lie}$ approaches the payoff the sender achieves in the traditional cheap talk model with these preliminaries, $\mathbb{E}[\theta]$.

The left panel of Figure 2 provides an intuition for Proposition 2($b$). It depicts the expected utility of the type-0 sender when she lies with $m = 1$ and $m = \theta_2$ as a function of $\sigma$, the probability she lies with $m = 1$. The type-0 sender's expected utility from lying with $m = 1$ is decreasing in $\sigma$ since as $\sigma$ increases, the receiver is more likely to verify $m = 1$ and takes a lower action following $m = 1$ when he does not.[16] On the other hand, as $\sigma$ increases, the receiver is less likely to verify $m = \theta_2$ and takes a higher action following $m = \theta_2$ when he does not. So the type-0 sender's expected utility from lying with $m = \theta_2$ is increasing in $\sigma$.[17] If $\beta \in (\underline{\beta}, \overline{\beta})$ these curves have a unique intersection point, denoted $\sigma^*$. If $\beta \geq \overline{\beta}$, lying with $m = 1$ becomes too attractive to the type-0 sender, and there is no longer a $\sigma^*$ such that she can be made indifferent between lying with $m = 1$ and $m = \theta_2$. And if $\beta \leq \underline{\beta}$, the receiver verifies both messages in the support of the sender's strategy with certainty. This case is depicted in the right panel of Figure 2. In particular, when $\beta \leq \tilde{\beta}$, there is an interval, $[\sigma_1, \sigma_0]$, such that if $\sigma \in [\sigma_1, \sigma_0]$, the receiver verifies both messages with certainty. In all IEs in this region of the parameter space, $\sigma^* \in [\sigma_1, \sigma_0]$.

There is one additional IE to consider, the IE described by Proposition 2($c$). In this IE, the sender tells the truth if $\theta \in \{\theta_2, 1\}$ and lies if $\theta = 0$. Yet, the type-0 sender never lies with $m = \theta_2$. This IE requires a knife-edge condition to be satisfied, $\beta = \overline{\beta}$, which ensures the expected utility of lying is equivalent to $\theta_2$. As such, even though the message $m = \theta_2$ is never verified, the type-0 sender has no incentive to deviate to sending it since it offers the same payoff she receives from lying with $m = 1$. Similarly, the type-$\theta_2$ sender does not have an incentive to deviate to lying since doing so yields the same payoff as telling the truth.

Previous theoretical work on cheap talk and costly verification finds that if the cost of

[16] In fact, as depicted in Figure 2, for $\beta$ sufficiently small and $\sigma$ sufficiently large ($\sigma > \sigma_1(\beta)$), the type-0 sender's expected utility from lying with $m = 1$ is zero since the receiver verifies the message with certainty.

[17] Although not depicted in Figure 2, for $\beta$ sufficiently small and $\sigma$ sufficiently small, the sender's expected utility from lying with $m = \frac{1}{2}$ is zero since the receiver verifies the message with certainty.

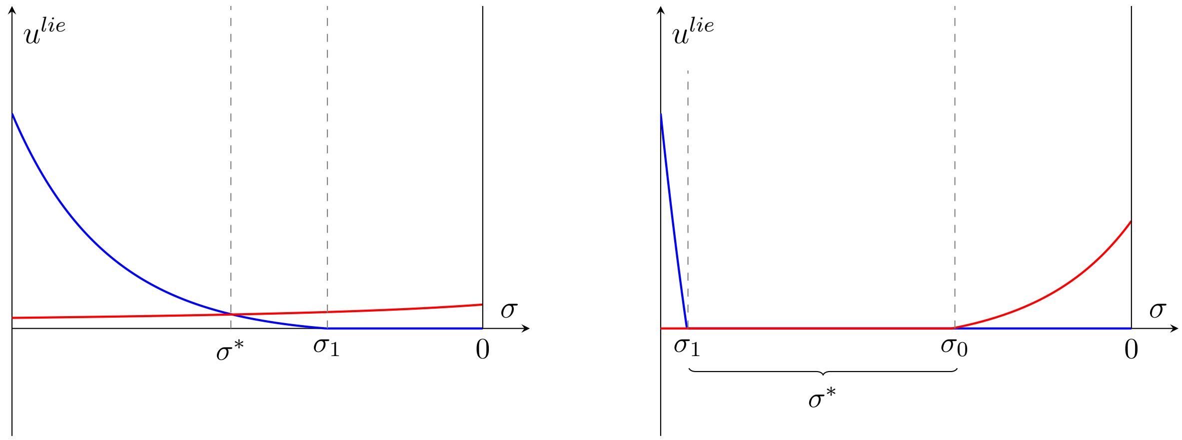


Figure 2: Left panel: Expected utility from lying with $m = 1$ and $m = \theta_2$ for a sender of type $\theta = 0$. Assumes $\beta = \frac{6}{25}$ and $\theta_2 = \frac{1}{3}$. Left panel: Expected utility from lying with $m = 1$ and $m = \theta_2$ for a sender of type $\theta = 0$. Assumes $\beta = \frac{1}{20}$ and $\theta_2 = \frac{1}{3}$.

verification is sufficiently high, verification does not occur in equilibrium, which destroys informative communication (e.g., Levkun, 2022; Sadakane and Tam, 2023). Proposition 2 shows that when the cost of verification is the receiver's private information, an influential equilibrium continues to exist even as the upper-bound of the cost of verification goes to infinity. Moreover, this IE is informative because the receiver's posterior is not constant on the equilibrium path.[18] However, the variation in the posterior comes from verification rather than from different messages conveying different information in equilibrium.

**The Effects of Verification** Let $\Gamma^{CT}$ denote the version of the three-state model that is identical except the receiver does not have the ability to verify. Put differently, $\Gamma^{CT}$ is the traditional cheap talk model. Perhaps unsurprisingly, the ability to verify the sender's message improves the receiver's expected utility relative to her expected utility in $\Gamma^{CT}$.[19]

**Proposition 3.** *In any IE, the receiver's expected utility is higher than his expected utility in* $\Gamma^{CT}$*.*

There are two reasons the receiver's expected utility is higher when he can verify the

[18] An equilibrium is informative if there are two beliefs on the equilibrium path, $\mu$ and $\mu'$, and $a(\mu) \neq a(\mu')$ (Sobel, 2013).

[19] In $\Gamma^{CT}$, no informative equilibria exists. As a result, the receiver chooses $a(m) = \mathbb{E}[\theta]$. See Footnote 7 for more information.

sender's message than in $\Gamma^{CT}$. The first is that verification has a direct *informational* effect, which is quantified by $\Lambda(m)$. The informational effect can be seen clearly in the IE described by Proposition 2($a$). In this IE, the type-1 sender is the only type of sender who tells the truth. But, verification allows the receiver to distinguish truthful messages sent by the type-1 sender from lies sent by the type-0 and type-$\theta_2$ sender. Moreover, while verification is costly, the fact that the cost is the receiver's private information means that, ex-ante, the informational effect outweighs the expected cost.[20]

Verification also increases the receiver's expected utility through an indirect *deterrence* effect. In equilibrium, verification generates an endogenous cost of lying for the type-$\theta_2$ sender. If she lies and is caught, she will be pooled with the type-0 sender. In some cases, this cost of lying deters the type-$\theta_2$ from lying.

The deterrence effect can be seen in the IE described by Proposition 2($b$).[21] In this IE, the type-$\theta_2$ sender tells the truth. Deviating to lying with $m = 1$ leads to a payoff of $a(1, \emptyset) > a(\theta_2, \emptyset)$ if the receiver does not verify. But when he does, the type-$\theta_2$ sender is punished by the receiver choosing an action that pools her with the type-0 senders who also lie. In this IE, this punishment is costly enough to deter the type-$\theta_2$ sender from lying.

In the IE described by Proposition 2($b$), the deterrence and informational effects of verification are both present. A natural conjecture is that these effects operate like substitutes: if verification deters lying, it reduces the informational value of verifying the sender's message. In the following proposition, I show this is not necessarily the case. To do so, I first introduce the concept of the expected informational effect of verification:

$$\mathbb{E}_{m\in M^*}[\Lambda(m)] = \Sigma_{m\in M^*}\rho(m)(Var(m|c=0) - (1-\phi(m))(Var(m|c=1, v=f))$$

[20]In contrast, in Sadakane and Tam (2023), the direct informational effect is completely offset by the cost of verification because the equilibrium is constructed such that when the receiver verifies, he is exactly indifferent between verifying and not verifying.

[21]This IE exists simultaneously with the IE in which verification only provides an informational effect, described by Proposition 2($a$), when $\beta \in (\tilde{\beta}, \overline{\beta})$

where $\rho(m)$ is the probability $m$ is sent on the equilibrium path.[22]

**Proposition 4.** *Suppose $\beta \in (\tilde{\beta}, \overline{\beta})$ and $\theta_2 > \overline{\theta}$ as defined in (13). For $\beta$ sufficiently close to $\overline{\beta}$, the expected informational effect of verification is greater in the IE where there is a deterrence effect than in the IE where there is not.*

An intuition for this result is as follows. When $\beta \in (\tilde{\beta}, \overline{\beta})$, two IEs exist. In one, verification only has an informational effect; in the other, it has informational and deterrence effects. Fix $\theta_2$. In the latter IE, increasing $\sigma^*$ has two effects: it increases the variance in the receiver's belief when he observes $m = 1$ and does not verify, and it increases the probability the sender lies with $m = 1$. Together, this means the informational effect of verification is increasing $\sigma^*$. Stated differently, the deterrence effect means the sender lies with more lies in equilibrium, and the expected informational effect of verification is higher when the sender lies relatively more often by pretending to be the highest type.

Now consider the IE in which verification only has an informational effect. This informational effect is decreasing in $\theta_2$. This is because as $\theta_2$ increases, the variance in the receiver's belief when he does not verify $m = 1$ is increasing, and the variance when he does is decreasing. Put differently, without deterrence, more sender types lie, and the informational effect of verification increases the more diverse this pool of liars is.

Consider the limiting case where $\sigma^* = 1$. If $\theta_2 > \overline{\theta}$, the expected informational effect of verification is greater in the IE where there is a deterrence effect than in the IE where there is not. Furthermore, for $\sigma^*$ sufficiently close to 1, this is still true. $\sigma^*$ is not a parameter; it is determined in equilibrium and depends on $\theta_2$ and $\beta$. However, for any $\theta$, if $\beta \to \overline{\beta}$, $\sigma^* \to 1$. Since $\sigma^*$ is continuous in $\beta$, which I show in the Appendix, if $\theta_2 > \overline{\theta}$, and if $\beta$ is sufficiently close to $\overline{\beta}$, the expected informational effect of verification is greater in the IE where there is a deterrence effect than in the IE where there is not.

[22]Recall, $\phi(m)$ is the probability $m$ is sent truthfully.

# 5 Conclusion

In this paper, I presented a model of cheap talk communication with costly verification. In the model, an informed sender communicates with a receiver with private information about the cost he will need to pay to verify the veracity of the sender's message. I show that any influential equilibrium of the model has a straightforward structure characterized by threshold: high-type senders tell the truth, and low-type senders lie by pretending to be high types. Except in a knife-edge case, any message the sender sends truthfully she also sends as a lie, and all messages in the support of the sender's strategy are verified with positive probability.

I then characterize all influential equilibria when there are three equally likely states and the cost to verify is drawn from a uniform distribution. In addition to showing how equilibria change depending on the upper bound of the cost the receiver pays to verify, this analysis illustrates how verification has two effects: an informational effect and a deterrence effect. I conclude by exploring these effects and showing that, in some cases, the presence of the deterrence effect increases the expected informational effect.

A key feature of the analysis of the three-state example is the observation that multiple equilibria exist within the same region of the parameter space, and there is variation within these equilibria related to the threshold that characterizes them about the extent to which verification has a deterrence effect. One might conjecture that this observation can be generalized to cases where there are more than three states. This conjecture provides an avenue for future research.

# 6 Appendix: Main Results

## 6.1 Proof of Proposition 1

*Proof.* Suppose an IE exists.

(i.) Claim: At least one message is sent as a lie.

*Proof.* Suppose not. Then all messages are sent truthfully. If the receiver observes $m$, he believes $\theta = m$, will not verify, and will choose $a = m$. But the sender of type $\theta < 1$ has a profitable deviation to $m = 1$. ■

(ii.) Claim: At least one message is sent truthfully.

*Proof.* Suppose not. Then all messages are sent as lies and the receiver never verifies. By Assumption 1, the sender's lying strategy cannot depend on her type, which means $\mathbb{E}[\theta|m]$ is constant for all $m$. But then $a^*(m, v)$ is constant on the equilibrium path. This violates the definition of an IE. ■

(iii.) Claim: The sender's expected utility from lying with message $m$ is constant for all lies in the support of the sender's strategy.

*Proof.* Let $M_f$ be the set of messages that are sent as lies on the equilibrium path. Suppose the claim is false. Then there exist lies $m \in M_f$ and $m' \in M_f$ such that the sender's expected utility from lying with $m$ is strictly higher than lying with $m'$. But then a liar who was supposed to lie with $m'$ could deviate to lying with $m$. Hence, this is not an IE. Denote the expected utility of lying $u^{lie}$. ■

(iv.) Claim: Each message $m$ such that $m \neq u^{lie}$ and $m$ is in the support of the sender's strategy is sent as a lie.

*Proof.* Let $M_t$ be the set of messages that are sent truthfully on the equilibrium path, and let $M^* = M_f \cup M_t$ be the set of messages in the support of the sender's strategy.

Suppose the claim is false. Then there exists some $m \in M^*$ such that $m$ is not in $M_f$ and $m \neq u^{lie}$. When the receiver observes $m$, he will not verify and will choose $a = m$.

Suppose $m > u^{lie}$. Then a liar can deviate from lying according to her equilibrium strategy to lying by reporting $m$. The message will believed by Assumption 2, so it is a profitable deviation.

Suppose $m < u^{lie}$. Then the type-$m$ sender can deviate from truthfully reporting $m$, in which case she receives a payoff of $u_S(m)$, to lying and get $u^{lie}$. This is a profitable deviation. ■

(v.) Claim: For any $\theta$ such that $\theta \neq u^{lie}$, the sender either lies with probability one or tells the truth with probability one.

*Proof.* Suppose not. Consider the type-$\theta$ where $\theta \neq u^{lie}$ who randomizes between lying and telling the truth. She must be indifferent between telling the truth and lying. By Assumption 1, all senders who lie must use the same strategy, so $m = \theta$ can only be sent truthfully. Hence, when the sender sends $m$, the receiver does not verify and chooses $a(m, \emptyset) = m$. It is immediate that the sender cannot be indifferent between telling the truth and lying. ■

(vi.) Claim: For any $\theta$ such that $\theta > u^{lie}$, the sender tells the truth with probability one.

*Proof.* Suppose not. Then there exists a $\theta$ such that $\theta > u^{lie}$ and the type-$\theta$ sender lies. Then $m = \theta$ is not in the support of the sender's strategy since the sender's strategy when lying is independent of $\theta$ and the type-$\theta$ cannot lie with $m = \theta$. Since $m$ is off the equilibrium path, when the sender lies and receives $u^{lie}$, she can deviate to the message $m$. This message will be believed by Assumption 2, and the receiver will choose $a(m, \emptyset) = \theta$. By assumption, this is larger than $u^{lie}$. ■

(vii.) Claim: For any $\theta$ such that $\theta < u^{lie}$, the sender lies with probability one.

*Proof.* Suppose not. Then there exists $\theta$ such that $\theta < u^{lie}$ and the type-$\theta$ sender tells the truth.

The type-$\theta$ sender's expected utility from telling the truth is $\pi^*(m)\theta + (1-\pi^*(m))a(m,\emptyset)$, where $\pi^*(m)$ is the probability the receiver verifies the message $m$ in equilibrium. Optimality implies $\pi^*(m)\theta + (1-\pi^*(m))a(m,\emptyset) \geq u^{lie}$, which requires $a(m,\emptyset) > u^{lie}$ since $\theta < u^{lie}$. However, $(vi.)$ shows that the sender tells the truth if $\theta > u^{lie}$. Hence $a(m,\emptyset) \leq \mathbb{E}[\theta|\theta \leq u^{lie}] < u^{lie}$. This is a contradiction. ■

(vii.) Claim: If there exists a $\theta$ such that $\theta = u^{lie}$, $m = \theta$ is only sent truthfully.

*Proof.* Suppose there exists a $\theta$ such that $\theta = u^{lie}$ and $m = \theta$ is sent as a lie.

Suppose further that $m = \theta$ is only sent as a lie. This implies the type-$\theta$ sender lies with probability one. But she cannot lie by truthfully reporting her type and Assumption 1 implies the sender's lying strategy is independent of her type.

Now suppose $m = \theta$ is sent truthfully. When $m = \theta$ is sent as a lie, it induces expected utility $\pi^*(m)a(m,f) + (1-\pi^*(m))a(m,\emptyset) = u^{lie}$. Since the sender lies if $\theta < u^{lie}$, $a(m,f) < u^{lie}$ and $a(m,\emptyset) < u^{lie}$. Hence, the required inequality cannot not hold. ■

**Corollary A.** *In any IE, the receiver verifies all $m > u^{lie}$.*

■

## 6.2 Proof of Proposition 2

*Proof.* **If direction:**

Proposition 1 implies that there are three possible structure to an IE: the receiver verifies all messages in the support of the sender's strategy with probability $\pi(m) \in (0,1)$, $\pi(m) = 1$ for all messages in the support of the sender's strategy, or there exists a $\theta$ such that $\theta = u^{lie}$ and $m = \theta$ is on path but only sent truthfully.

$\underline{\pi(m) \in (0,1) \text{ for } m \text{ on the path:}}$

Suppose an IE exists where the receiver verifies each message in the support of the sender's strategy probability with probability $\pi(m) \in (0,1)$. Proposition 1 implies there are two possible structures to this IE.

First, suppose an IE exists where the sender lies when $\theta \in \{0, \theta_2\}$ by reporting $m = 1$, and tells the truth when $\theta = 1$. The type-0 sender and type-$\theta_2$ sender's expected utility from lying with $m = 1$ is:

$$\min\left\{\frac{(2-\theta_2)^2}{18\beta}, 1\right\} u_S\left(\frac{\theta_2}{2}\right) + \max\left\{1 - \frac{(2-\theta_2)^2}{18\beta}, 0\right\} u_S\left(\frac{\theta_2+1}{3}\right). \tag{3}$$

This is a piecewise function that is constant for $\beta \in (0, \frac{(2-\theta_2)^2}{18})$ since $\min\{\frac{(2-\theta_2)^2}{18\beta}\} = 1$, and is strictly increasing for $\beta > \frac{(2-\theta_2)^2}{18}$:

$$\frac{\partial(3)}{\partial\beta} = -\frac{(2-\theta_2)^2}{18\beta^2} u_S\left(\frac{\theta_2}{2}\right) + \frac{(2-\theta_2)^2}{18\beta^2} u_S\left(\frac{\theta_2+1}{3}\right) > 0$$

since $\frac{\theta_2+1}{3} > \frac{\theta_2}{2}$.

Neither the type-0 sender nor the type-$\theta_2$ sender have have a profitable deviation from lying as long as (3) $\geq \theta_2$ since $m = \theta_2$ is off the equilibrium path and can be obtained by deviating due to Assumption 2. Since (3) is weakly increasing in $\beta$, this condition is satisfied as long as $\beta \geq \tilde{\beta}$, where

$$\tilde{\beta} \equiv \frac{(2-\theta_2)^3}{36(1-2\theta_2)}. \tag{4}$$

This is defined for all $\theta_2 \in (0, \frac{1}{2})$.

The type-1 sender never has a profitable deviation from telling the truth since:

$$\min\left\{\frac{(2-\theta_2)^2}{18\beta}, 1\right\} u_S(1) + \max\left\{1 - \frac{(2-\theta_2)^2}{18\beta}, 0\right\} u_S\left(\frac{\theta_2+1}{3}\right) > (3)$$

Then, since $\frac{(2-\theta_2)^3}{36(1-2\theta_2)} > \frac{(2-\theta)^2}{18}$ for all $\theta_2 \in (0, \frac{1}{2})$, this IE exists for all $\beta \geq \tilde{\beta}$.

Second, suppose an IE exists where the sender tells the truth when $\theta \in \{\theta_2, 1\}$, and lies when $\theta = 0$ by randomizing over $m \in \{\theta_2, 1\}$. Proposition 1 implies that in this IE, $\sigma^* = \sigma$ solves

$$\max\left\{\left(1 - \frac{\sigma}{\beta(1+\sigma)^2}\right), 0\right\} u_S\left(\frac{1}{1+\sigma}\right) = \max\left\{\left(1 - \frac{(1-\sigma)\theta_2^2}{\beta(2-\sigma)^2}\right), 0\right\} u_S\left(\frac{\theta_2}{(2-\sigma)}\right). \quad (5)$$

The left-hand side of (5) is a piecewise function that is strictly decreasing in $\sigma$ for $\sigma \in (0,1)$ when $\beta > \frac{\sigma}{(1+\sigma)^2}$:

$$\frac{\partial}{\partial\sigma} \max\left\{\left(1 - \frac{\sigma}{\beta(1+\sigma)^2}\right), 0\right\} u_s\left(\frac{1}{1+\sigma}\right) = \\ -\frac{1-\sigma}{\beta(1+\sigma)^3} u_S\left(\frac{1}{1+\sigma}\right) + \left(1 - \frac{\sigma}{\beta(1+\sigma)^2}\right) \frac{\partial}{\partial\sigma} u_S\left(\frac{1}{1+\sigma}\right) < 0,$$

since $u_S(a)$ is increasing in $a$. Moreover, (5) is constant for $\sigma \in (0,1)$ when $\beta < \frac{\sigma}{(1+\sigma)^2}$ since $\max\{(1 - \frac{\sigma}{\beta(1+\sigma)^2}), 0\} = 0$, $\to u_S(1)$ when $\sigma \to 0$, and $\to u_S(\max\{\frac{4\beta-1}{8\beta}, 0\})$ when $\sigma \to 1$.

The right-hand side of (5) is a piecewise function that is strictly increasing in $\sigma$ for $\sigma \in (0,1)$ when $\beta > \frac{(1-\sigma)\theta_2^2}{(2-\sigma)^2}$:

$$\frac{\partial}{\partial\sigma} \max\left\{\left(1 - \frac{(1-\sigma)\theta_2^2}{\beta(2-\sigma)^2}\right), 0\right\} u_S\left(\frac{\theta_2}{(2-\sigma)}\right) = \\ \frac{\sigma\theta_2^2}{\beta(2-\sigma)^3} u_S\left(\frac{\theta_2}{2-\sigma}\right) + \left(1 - \frac{(1-\sigma)\theta_2^2}{\beta(2-\sigma)^2}\right) \frac{\partial}{\partial\sigma} u_S\left(\frac{\theta_2}{2-\sigma}\right) > 0.$$

since $u_S(a)$ is increasing in $a$. Moreover, (5) is constant for $\sigma \in (0,1)$ when $\beta < \frac{(1-\sigma)\theta_2^2}{(2-\sigma)^2}$, which implies $\max\{(1 - \frac{(1-\sigma)\theta_2^2}{\beta(2-\sigma)^2}, 0\} = 0$, $\to u_S(\max\{\frac{(4\beta-\theta_2^2)\theta_2}{8\beta}, 0\})$ when $\sigma \to 0$, and $\to u_S(\theta_2)$ when $\sigma \to 1$.

Hence, there is at least one $\sigma^* \in (0,1)$ such that(5) is satisfied as long as $u_S(\frac{4\beta-1}{8\beta}) > u_S(\theta_2)$, which is satisfied if and only if $\beta < \overline{\beta}$, where

$$\overline{\beta} \equiv \frac{1}{4(1-2\theta_2)}. \quad (6)$$

Moreover, given $\sigma^*$, the receiver verifies both messages in the support of the sender's strategy with probability $\pi(m) \in (0,1)$ unless

$$\frac{1-\sqrt{1-4\beta}-2\beta}{2\beta} < \frac{4\beta-\theta_2^2+\sqrt{-4\beta\theta_2^2+\theta_2^4}}{2\beta} \tag{7}$$

where the left-hand side of (7) is the value of $\sigma$ such that $\beta = \frac{\sigma}{(1+\sigma)^2}$ and the right-hand side of (7) is the value of $\sigma$ such that $\beta = \frac{(1-\sigma)\theta_2^2}{(2-\sigma)^2}$.

The left-hand side of (7) is defined for $\beta < \frac{1}{4}$, is increasing in $\beta$:

$$\frac{\partial}{\partial\beta}\frac{1-\sqrt{1-4\beta}-2\beta}{2\beta} = \frac{1-\sqrt{1-4\beta}-2\beta}{2\beta^2\sqrt{1-4\beta}} > 0,$$

$\to 1$ when $\beta \to \frac{1}{4}$, and $\to 0$ when $\beta \to 0$.

The right-hand side of (7) is defined for $\beta < \frac{\theta_2^2}{4}$, is decreasing in $\beta$ over this range:

$$\frac{\partial}{\partial\beta}\frac{\sqrt{1-36\beta}+36\beta-1}{18\beta} = \frac{2\beta\theta_2^2-\theta_2^4+\theta_2^2\sqrt{-4\beta\theta_2^2+\theta_2^4}}{2\beta^2\sqrt{-4\beta\theta_2^2+\theta_2^4}} < 0$$

$\to 0$ when $\beta \to \frac{\theta_2^2}{4}$, and $\to 1$ when $\beta \to 0$. Hence, for $\beta > \underline{\beta}$, where $\underline{\beta}$ is the value of $\beta$ such that

$$\frac{1-\sqrt{1-4\beta}-2\beta}{2\beta} = \frac{4\beta-\theta_2^2+\sqrt{-4\beta\theta_2^2+\theta_2^4}}{2\beta}, \tag{8}$$

(7) is not satisfied and there exists a unique $\sigma^*$ such that (5) is satisfied.

It is clear that a type-0 sender does not have a profitable deviation to truthfully reporting $m = 0$ as Assumption 2 implies the message will be believed, yielding her a payoff of $u_S(0)$, which is lower than the payoff she receives from lying, $u^{lie}$, which is positive. Additionally, neither the type-$\theta_2$ nor the type-1 sender have a profitable deviation to lying because when the receiver verifies and learns the message was true, he chooses $a = 1 > 0$ or $a = \theta_2 > 0$.

Additionally, note that because $\underline{\beta} < \frac{\theta_2^2}{4}$ and $\tilde{\beta} = \frac{1}{4(1-2\theta_2)}$, $\underline{\beta} < \tilde{\beta}$ for all $\theta_2$.

$\underline{\pi(m) = 1 \text{ for } m \text{ on the path}}$:

Now, suppose an IE exists where the receiver verifies each message in the support of the sender's strategy with probability $\pi(m) = 1$. Proposition 1 implies there are two possible IEs.

First, suppose an IE exists where the sender lies when $\theta \in \{0, \theta_2\}$ by reporting $m = 1$, and tells the truth when $\theta = 1$. Since the receiver verifies $m = 1$ with probability one, the sender of type ($\theta = \theta_2$ receives a payoff of $u_S(\frac{\theta_2}{2})$, which is lower than the payoff she would get from deviating to $m = \theta_2$.

Second, suppose there is an IE where the sender of type $\theta \in \{\theta_2, 1\}$ tells the truth and the sender of type $\theta = 0$ lies by randomizing between $m \in \{\theta_2, 1\}$. The type-0 sender does not have an incentive to deviate because her payoff is $u_S(0)$ on the path and $u_S(0)$ if she deviates to $m = 0$. Moreover, neither the type-$\theta_2$ sender nor the type-1 sender have an incentive to deviate from telling the truth because doing so will lead to a payoff of $u_S(0)$ which is lower than the payoff on the equilibrium path. Given the previous analysis, this IE exists if there is a $\sigma^*$ such that $\frac{\sigma^*}{(1+\sigma^*)^2} \geq \beta$ and $\frac{(1-\sigma^*)\theta_2^2}{(2-\sigma^*)^2} \geq \beta$. That is, this IE exists for $\sigma$

$$\sigma \in [\sigma_1, \sigma_0], \tag{9}$$

where $\sigma_1 = \frac{-1-\sqrt{1-16\beta}+16\beta}{8\beta}$ and $\sigma_0 = \frac{4\beta-\theta_2^2+\sqrt{-4\beta\theta_2^2+\theta_2^4}}{2\beta}$. Previously, it was shown that this interval is not empty as long as $\beta \leq \underline{\beta}$. Moreover, the previous discussion implies that $\frac{4\beta-\theta_2^2+\sqrt{-4\beta\theta_2^2+\theta_2^4}}{2\beta} < \frac{1-\sqrt{1-4\beta}-2\beta}{2\beta}$ if $\beta < \tilde{\beta}$.

$\underline{\theta = u^{lie} \text{ and } m = \theta \text{ is sent truthfully}}$:

Finally, suppose an IE exists where there is a $\theta$ such that $\theta = u^{lie}$ and $m = \theta$ is on path and only sent truthfully. This implies $\pi(\theta) = 0$. Proposition 2 implies that when $\Theta = \{0, \theta_2, 1\}$, the only case where this IE might exists is if $u^{lie} = \theta_2$. In such an IE, the type-0 sender lies, and her expected utility is $\frac{4\beta-1}{8\beta}$. Hence, such an IE exists if $\theta_2 = \frac{4\beta-1}{8\beta}$. Rearranging, this IE exists if $\beta = \overline{\beta}$.

**Only if direction:**

Suppose $\beta \geq \tilde{\beta}$. It is immediate from the previous analysis that the type-0 and type-$\theta_2$ sender prefer lying with $m = 1$ to truthfully reporting the state or deviating to an off the path message. Moreover, the type-1 sender prefers truthfully reporting the state to lying or deviating. This implies an IE exists in which the players strategies are as described in Proposition 2($a$).

Suppose next that $\beta \in (\underline{\beta}, \overline{\beta})$. It is immediate from the previous analysis that there is a unique $\sigma^*$ satisfying (5). Moreover, given this $\sigma^*$, the type-0 sender prefers to lie than deviating to $m = 0$ and the type-1 sender and type-$\theta_2$ sender prefer to tell the truth to lying or deviating off path. This implies an IE exists in which the players strategies are as described in Proposition 2($b$).

Suppose that $\beta = \overline{\beta}$. If the type-0 sender lies and reports $m = 1$, $u^{lie} = \theta_2$. Hence, an IE exists in which the players strategies are as described in Proposition 2($c$).

Finally, suppose $\beta \leq \underline{\beta}$. It is immediate from the previous analysis that there is a continuum of $\sigma^*$ such that the type-0 sender is weakly prefers to lie and the type-$\theta_2$ and type-1 sender prefer to tell the truth than deviate off path or lie. Hence, an IE exists in which the players strategies are as described in Proposition 2($d$). ■

## 6.3 Proof of Proposition 3

*Proof.* In $\Gamma^{CT}$, the receiver's expected utility is:

$$-\frac{2(1-\theta_2+\theta_2^2)}{9}. \tag{10}$$

In the IE described by Proposition 2($a$), the receiver's expected utility is $u_R^a = \frac{(\theta_2-2)^4}{648\beta} - \frac{2(1-\theta_2+\theta_2^2)}{9}$, which is clearly larger than (10).

In the IE described by Proposition 2($b$), due to the optimality of the receiver's best response, his expected utility, $u_R^b$, is bounded below by $\overline{u}_R^b = -\frac{\sigma}{3(1+\sigma)} - \frac{(1-\sigma)\theta_2^2}{3(2-\sigma)}$, which is his expected utility if he never verifies. It can be shown through algebraic manipulation that $\overline{u}_R^b >$ (10) for all $\sigma$ and $\theta_2$, which implies $u_R^b >$ (10).

In the IE described by 2($c$), the receiver's expected utility is $u_R^c = -\frac{\theta^2}{3}$, which is larger than (10).

And in the IEs described by 2($d$), the receiver's expected utility is $\int_0^1(-\beta\kappa)d\kappa = -\frac{\beta}{2}$. This type of IE exists for $\beta \leq \underline{\beta}$, and $\underline{\beta} < \frac{1}{4}$ by its definition. Hence, $-\frac{\beta}{2} >$ (10) for the $\beta$ such that this type of IE exists. ■

## 6.4 Proof of Proposition 4

To prove Proposition 4, I begin with the following lemma.

**Lemma A.** *Consider the IE described by Proposition 4. If $\beta \in (\underline{\beta}, \overline{\beta})$, $\sigma^*(\beta)$ is continuous.*

*Proof.* Abusing notation, define an implicit function $h(\sigma, \beta) = f(\sigma, \beta) - g(\sigma, \beta)$, where $f(\sigma, \beta)$ and $g(\sigma, \beta)$ are the type-0 sender's expected utility from lying with $m = 1$ and $m = \theta_2$ respectively. Proposition 2 implies that for every $\beta \in (\underline{\beta}, \overline{\beta})$ there exists a unique $\sigma^*(\beta)$ such that $h(\sigma^*(\beta), \beta) = 0$. Proposition 2 also shows that $f(\sigma, \beta)$ and $g(\sigma, \beta)$ are continuously differentiable in $\beta$ and $\sigma$. Hence, $h(\sigma, \beta)$ is too. Applying the implicit function theorem, for an arbitrary $\beta \in (\underline{\beta}, \overline{\beta})$, there exists a unique, differentiable function $\phi(\beta)$ such that in the neighborhood of $\beta$, $h(\phi(\beta), \beta) = 0$ as long as $\frac{\partial f}{\partial \sigma}(\phi(\beta), \beta) - \frac{\partial g}{\partial \sigma}(\phi(\beta), \beta) \neq 0$. In the proof of Proposition 2, I showed that $\frac{\partial f}{\partial \sigma} < 0$ and $\frac{\partial g}{\partial \sigma} > 0$, ensuring this condition is satisfied. Then, because there is a unique $\sigma^*(\beta)$ and unique $\phi(\beta)$ for all $\beta \in (\underline{\beta}, \overline{\beta})$, $\sigma^*(\beta) = \phi(\beta)$. Moreover, the local continuity of $\phi(\beta)$ on a connected space implies global continuity on the space. ■

I now prove Proposition 4.

*Proof.* Suppose $\beta \in (\tilde{\beta}, \overline{\beta})$. Then an IE exists where there is a deterrence benefit from verification (described by Proposition 2($a$)) and an IE exists where there is not (described by Proposition 2($b$)). In the latter, the expected informational effect of verification is

$$\frac{(2-\theta)^2}{18}. \tag{11}$$

In the former, the expected informational effect of verification is

$$\frac{(2-\sigma^*)\sigma^* + \theta_2^2(1-\sigma^{*2})}{3(2-\sigma^*)(1+\sigma^*)}. \tag{12}$$

(12) is increasing in $\sigma^*$:

$$\frac{\partial(12)}{\partial\sigma^*} = \frac{1}{3}\left(\frac{1}{(1+\sigma^*)^2} - \frac{\theta_2^2}{(2-\sigma^*)^2}\right) > 0$$

Moreover, for $\sigma^* = 1$, (12) $= \frac{1}{6}$. Hence, if

$$\begin{aligned} (11) &< \frac{1}{6} \\ \Leftrightarrow \overline{\theta}_2 \equiv 2 - \sqrt{3} &< \theta_2, \end{aligned} \tag{13}$$

then for $\sigma^*$ sufficiently close to one, (12) $>$ (11).

Fix $\theta_2 > \overline{\theta}_2$. Recall from the proof of Proposition 2 that for all $\beta \in (\underline{\beta}, \overline{\beta})$, there is a unique $\sigma^*$ that solves (5). Taking the limit of $f(\sigma, \theta)$ and $g(\sigma, \theta)$ as $\beta \to \overline{\beta}$, it is clear that (5) is uniquely solved by $\sigma^* = 1$ for $\beta = \overline{\beta}$. Proposition 2 implies that for all $\beta \in (\tilde{\beta}, \overline{\beta})$, in the IE described by Proposition 2($b$) $\sigma^* \in (0, 1)$. Moreover, Lemma A implies $\sigma^*(\beta)$ is continuous. Hence, for $\beta$ sufficiently close to $\overline{\beta}$, (12) $>$ (11). ■

# 7 Supplementary Appendix: Additional Results

## 7.1 Existence of an IE

**Proposition A.** *Define $\check{\beta}$ and $\dot{\beta}$ as in (14) and (15). An IE exists if at least one of the following is satisfied:*

*(a.)* $\mathbb{E}[\theta] \geq \theta_{N-1}$ *and* $\beta \geq \check{\beta}$.

*(b.)* $\beta \leq \dot{\beta}$.

*Proof.* Consider a strategy for the sender where she lies if $\theta \in \{\theta_1, ..., \theta_{N-1}\}$ by reporting $m = \theta_N$ and tells the truth if $\theta = \theta_N$. Then when the sender lies, her expected utility is $\pi(\theta_N)\mathbb{E}[\theta|\theta \in \{\theta_1, \theta_{N-1}\}] + (1 - \pi(\theta_N))\mathbb{E}[\theta]$, where

$$\pi(\theta_N) = G\left(\frac{Var(\theta_N|c = 0) - (1 - \phi(m)Var(\theta_n|v = f)}{\beta}\right).$$

The only effect of $\beta$ on $\pi(\theta_N)$ is through the denominator, which means $\pi(\theta_N)$ is decreasing in $\beta$. Hence, $\mathbb{E}[\theta] > \theta_{N-1}$, there is a unique $\beta$ such that

$$\pi(\theta_N)\mathbb{E}[\theta|\theta \in \{\theta_1, \theta_{N-1}\}] + (1 - \pi(\theta_N))\mathbb{E}[\theta] = \theta_{N-1}. \tag{14}$$

Denote this $\beta$ as $\check{\beta}$. For all $\beta \geq \check{\beta}$, this IE exists.

Now consider a strategy for the sender where she lies if $\theta = \theta_1$ and tells the truth if $\theta > \theta_1$. In particular, when she lies, she uses a strategy where the probability she lies with $m = \theta_i \in \{\theta_2, ..., \theta_N\}$ is $\sigma_i \in (0, 1)$. A vector of lying probabilities, $\boldsymbol{\sigma}$ induces a vector of probabilities of verification, $\boldsymbol{\pi}$, where, for a particular $\pi(m) \in \boldsymbol{\pi}$, $\pi(m) = G(\frac{Var(m|c=0)}{\beta})$. Fix $\boldsymbol{\sigma}$. The only effect of $\beta$ on $\pi(m)$ is through the denominator. So there exists a $\beta$ such that

$$1 = \min(\pi(m) \in \boldsymbol{\pi}). \tag{15}$$

Denote this $\beta$ by $\dot{\beta}$. Moreover, for all $\beta \leq \dot{\beta}$, the receiver verifies each $m$ in the support of the sender's strategy with certainty. ■

## References


Balbuzanov, I. (2019): "Lies and consequences," *International Journal of Game Theory*, 48, 1203–1240.

Ball, I. and X. Gao (2025): "Checking cheap talk," *Working Paper*.

Dziuda, W. and C. Salas (2018): "Communication with detectable deceit," *Working Paper*.

Ederer, F. and W. Min (2022): "Bayesian Persuasion with Lie Detection," .

Gottfried, J. A., B. W. Hardy, K. M. Winneg, and K. H. Jamieson (2013): "Did fact checking matter in the 2012 presidential campaign?" *American Behavioral Scientist*, 57, 1558–1567.

Guo, Y. and E. Shmaya (2019): "The interval structure of optimal disclosure," *Econometrica*, 87, 653–675.

Hart, S., I. Kremer, and M. Perry (2017): "Evidence games: Truth and commitment," *American Economic Review*, 107, 690–713.

Holm, H. J. (2010): "Truth and lie detection in bluffing," *Journal of Economic Behavior & Organization*, 76, 318–324.

Kartik, N., M. Ottaviani, and F. Squintani (2007): "Credulity, lies, and costly talk," *Journal of Economic theory*, 134, 93–116.

Levkun, A. (2022): "Communication with strategic fact-checking," Tech. rep., mimeo. University of California, San Diego.

Lim, C. (2018): "Can Fact-checking Prevent Politicians from Lying?" *Working Paper*.

Nguyen, A. and T. Y. Tan (2021): "Bayesian persuasion with costly messages," *Journal of Economic Theory*, 193, 105212.

Nyhan, B. and J. Reifler (2010): "When corrections fail: The persistence of political misperceptions," *Political Behavior*, 32, 303–330.

——— (2015): "The effect of fact-checking on elites: A field experiment on US state legislators," *American Journal of Political Science*, 59, 628–640.

Pence, M. (2020): "Mike Pence, Dr. Fauci Coronavirus Task Force Press Conference June 26," https://www.rev.com/transcripts/mike-pence-dr-fauci-coronavirus-task-force-press-conference-june-26.

Pingree, R. J., D. Brossard, and D. M. McLeod (2014): "Effects of journalistic adjudication on factual beliefs, news evaluations, information seeking, and epistemic political efficacy," *Mass Communication and Society*, 17, 615–638.

Qui, L. (2020): "As Cases Surge, Pence Misleads on Coronavirus Pandemic," *The New York Times*, https://www.nytimes.com/2020/06/26/us/politics/coronavirus-pence-fact-check.html.

Sadakane, H. and Y. C. Tam (2023): "Cheap talk and Lie detection," *Working Paper*.

Sobel, J. (2013): "Giving and receiving advice," *Advances in Economics and Econometrics*, 1, 305–341.

——— (2020): "Lying and deception in games," *Journal of Political Economy*, 128, 907–947.

Venkatesh, R. S., S. Roy, and S. Pramanik (2025): "Bayesian Persuasion and Costly Verification: A model of two-sided commitment," *Working Paper*.

Weeks, B. E. (2015): "Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation," *Journal of Communication*, 65, 699–719.

Weeks, B. E. and R. K. Garrett (2014): "Electoral consequences of political rumors: Motivated reasoning, candidate rumors, and vote choice during the 2008 US presidential election," *International Journal of Public Opinion Research*, 26, 401–422.