Cheap talk and Lie detection*

Hitoshi Sadakane† Yin Chi (Terry) Tam‡§

June 14, 2022

Abstract

This paper analyzes strategic interactions between cheap talk and lie detection and studies the optimal equilibrium for costly lie detection and its effectiveness. An informed sender wants to persuade an uninformed receiver to take high actions, but the receiver wants to match the action with the true state. The sender makes a claim about the state, and the receiver decides whether to incur a cost to inspect the truthfulness of the claim. We show that the receiver-optimal equilibrium partitions the state space into three intervals. Types in the top interval make precise and truthful claims about the state, which are mimicked by types in the bottom interval and randomly inspected. Types in the middle interval make a vague claim that is never inspected. We show that lie detection is more beneficial to the receiver than state verification because it provides incentives for moderate and high types to be truthful.

JEL classification: C72; D82; D83

*We are grateful to Sergei Severinov, Li Hao, Vitor Farinha Luz, Michael Peters, Wei Li, Oliver Hart, Shintaro Miura, Takashi Shimizu, Takakazu Honryo, Kohei Kawamura, Ming Li, and seminar participants at University of British Columbia, East China University of Science and Technology, The University of Tokyo, Kyoto University, Kobe University, Hosei University, Osaka University of Economics, Contract Theory Workshop, Communication and Persuasion Workshop, and SWEO for helpful comments, discussions, and feedback. This draft is based on Tam (2019) previously circulated as "Lying and Lie-detecting" and Sadakane (2020) previously circulated as "Cheap talk and Fact-checking." Sadakane gratefully acknowledge the financial support from the Grants-in-Aid for Scientific Research (17H06778, 19H01471). Tam is sponsored by the Fundamental Research Funds for the Central Universities for this project.

†Institute of Economic Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan; email: sadakane.hitoshi.6c@kyoto-u.ac.jp

‡Oliver Hart Research Center of Contracts and Governance, East China University of Science and Technology, Shanghai, 200237, China; email: terrytamhk2012@gmail.com

§Corresponding author.

Keywords: Strategic Lie Detection; Cheap Talk; Strategic Information Transmission; Incomplete Information

1 Introduction

This paper introduces a theory of lying in cheap-talk communication where the receiver has access to a costly lie-detecting technology. To this end, we enrich the cheap-talk model, developed by Crawford and Sobel (1982)(hereafter, the CS model). In the CS model, a sender who knows the state $\theta \in [0,1]$ that a receiver would like to know can send a cheap-talk message to the receiver. After receiving the message, the receiver chooses an action. Unlike in the CS model, we consider a scenario in which (i) the sender's preference is state independent, (ii) the sender can send a message with a literal meaning about the state (e.g., "the realized state is θ"), and (iii) the receiver can verify whether the sender's message is a lie by conducting a costly inspection.[1] Throughout this paper, we refer a "claim" to as the literal meaning carried by the sender's message.

In this paper, we investigate whether and under what conditions is lie-detecting technology helpful in improving the welfare of the uninformed receiver. We first show that there exists an equilibrium where lying and lie detection occur if and only if inspection cost is sufficiently low and prior expectation of the state is not too high. Further, we characterize the receiver's optimal equilibrium and investigate the effectiveness of the strategic lie detection in improving the quality of communication.

The intuition behind the first result is simple. The threshold for prior expectation increases as inspection cost decreases and converges to the upper bound of the state space as inspection cost goes to zero. Intuitively, this result comes from the conflict between the sender's incentive to lie and the receiver's incentive to inspect. Liars aim to convince the receiver that they are better than the average (prior expectation) when they get away with the lie. If the prior expectation is too high, this can happen only when a small number of liars mimic a large number of truth-tellers, but then the message is not worth inspecting because the sender is too likely to be truthful. This result echoes a common perception that lie detection is effective when the sender is suspicious, in the sense that there is a substantial difference between the receiver's prior belief and the belief preferred by the sender. For example, the police usually conduct an interrogation only if they believe that the suspect is likely to have committed a crime. When the sender is likely to be "innocent," there is no cost-effective way to

[1] Sobel (2020) establishes a general framework of lying with various applications. Our model adopts the same definition of lying as in Sobel.

separate lies from truths.

Assuming the state is uniformly distributed, we show that the optimal equilibrium is characterized by three intervals, which partition the state space. The sender makes truthful claims when the true state is in the high interval (truthful types), and he lies and mimics one of the high claims when the true state is in the low interval (lying types). These high claims are randomly inspected. In the fear of being caught lying and perceived as low types, the sender in the intermediate interval (moderate types) is deterred from mimicking the high claims. These moderate types pool at a vague yet truthful claim which is not inspected by the receiver. It is optimal for the receiver to give the sender an option of being vague because precise claims require inspections to sustain, while moderate types are not distant enough from each other to justify the cost of inspection. Technically speaking, it is always optimal to pool an interval of types to a single message and leave it uninspected because the conditional variance of a small enough interval is lower than the inspection cost.

The key to the equilibrium construction is the mutual dependence between the sender's message strategy and the receiver's lie detection. In the equilibrium, the message strategies of the lying types add noise to the truthful types' messages. Therefore, the receiver who receives a message "the realized state is θ" cannot tell whether this message originates from the truthful type. Hence, the receiver has an incentive to inspect for removing the noise added by the lying types. The stochastic lie detection by the receiver makes the sender of lying types and moderate types indifferent between the outcome associated with the uninspected message and the lotterys over the outcomes induced by sending a lying message. From the truthful types' viewpoint, mimicking other truth-telling types induces the receiver's misunderstanding, which can eventually result in an unfavorable decision. This implies that the lying costs that deter truth-telling types from conveying misinformation are endogenously determined in the equilibrium.

The important implication in this paper is that lie-detecting technology improves the receiver's welfare only if lying occurs in equilibrium. If the sender never lies, the receiver has no incentive to inspect the sender's claim; if there is no inspection, babbling is the only equilibrium as the sender and receiver share no common interest. The receiver's benefit from lie detection can be decomposed into two components. First, lie detection generates a direct information value by distinguishing liars from truth-tellers which generally provides information about the true state. Second, the sender might stay honest in the fear of being caught lying. Therefore, the possibility of lie detection creates a threat that deters potential liars and facilitates information transmission. This is called the indirect deterrent

effect. Our results show that under the optimal lie-detecting equilibrium, the direct information value of inspection is completely offset by the cost of inspection. Improvement of the receiver's ex-ante payoff is driven by the deterrent effect: the receiver is able to elicit information from the sender due to the credible threats of lie detection. This is perhaps surprising as one might expect an optimal equilibrium should allow the receiver to acquire as much information from lie detection as possible. It turns out that an excessive amount of information acquired from inspection indicates that the equilibrium induces the sender to lie too often, and costly inspection takes place more frequently than the optimal equilibrium. This suggests that lie-detecting technology better serves as a means of deterrence than a means of information acquisition.

Our framework reflects the fundamental signal structure of lie-detecting technology broadly used in our daily lives, e.g., fact-checking. The motivating example illustrating the signal structure is as follows: Consider a situation in which a politician said "the Intergovernmental Panel on Climate Change (IPCC) has reported that limiting global warming to 1.5°C implies reaching net zero CO2 emissions globally around 2050 and concurrent deep reductions in emissions of non-CO2 forcers, particularly methane." In such a case, we can check a series of reports issued by IPCC and verify whether the politician is telling the truth.[2] However, reading through reports is costly. In addition, without fact-checking, the politician's statement is just a cheap-talk message, although it has literal meaning.

Under more general distributions, it is possible that inspected vague messages exist in the optimal equilibrium without the restriction on randomizing inspection. A vague message could pool moderate types and high types into the same group with a sufficiently high posterior such that inspecting the message is credible. It enables the separation of the moderate types from the lower types which would have been non-credible if the moderate types had sent precise messages. An insight from this observation is that when the market is flooded with low-quality products, having a quality standard that covers a wider range of high-quality types might be beneficial to the consumers, because it provides sufficient incentive for the consumers to verify the standard and separate a wider range of high-quality products from the low-quality products.

We study the effect of inspection technology on the receiver's welfare by comparing lie-detecting technology with state-verifying technology. There are substantial differences between lie detection and state verification. Under state-verifying technology, an inspection reveals the true state of the world.

[2]This statement is true. See, IPCC SR15 2018. (https://www.ipcc.ch/sr15/). As you will see in the link, checking the fact is a costly effort.

There will be no uncertainty upon inspection. Under lie-detecting technology, an inspection returns a binary signal on the truthfulness of the sender's claim. Information learned from an inspection is endogenously determined by the claim made by the sender. Practically, state verification is a hard skill which requires the receiver to be able to acquire knowledge about the true state, which might not be feasible in some situations. In the example of IPCC we discussed above, the effort of going through a series of reports reveals whether the politician's statement is true. State verification, i.e., verifying whether the content of statements is coherent with the state of the world, is another story. Also, there might not be any objective evidence in the crime scene that provides further information about whether a suspect has committed the crime. In this regard, lie detection can be a soft skill. A competent detective might be able to spot a lie told by the suspect using various interrogation tactics. Studies in psychology and cognitive science have shown possibilities of detecting lies using methods such as asking questions that raise cognitive load (Vrij et al. (2011)), measuring brain activities (Christ et al. (2008)) and reading micro-expressions (Porter and Ten Brinke (2006)), with nearly 70 percent accuracy (Hartwig and Bond (2014)) and 85 percent accuracy for trained interviewers (Hartwig et al. (2006)).

Even if state verification is feasible, lie-detecting technology can yield a higher benefit to the receiver. Assuming the same unit cost for the two technologies, we show that the receiver's welfare is higher under the optimal lie-detecting equilibrium compared with optimal state-verifying equilibrium. This is because revealing the true state upon inspection removes any strategic uncertainty that can serve as a threat of punishment to potential deviators. Since state verification leads to an accurate assessment of the true state, there is no credible punishment for the liar; thus the sender always has the incentive to exaggerate the state to "try his luck." As a result, the deterrence effect is eliminated under state-verifying technology, and there will not be any informative communication. This result sheds light on the optimal approaches of fact-checking as a tool to combat misinformation in politics. The internet has enabled the public to verify politicians' claims more easily using fact-checking websites such as FactCheck.org and PolitiFact. A question regarding the socially desirable mission of these organizations is whether they should focus on presenting verdicts on politicians' statements (lie detection) or educating the public about policy-related issues (state verification). The latter is more informative as verdicts on politicians' statements can be derived from knowledge in policy-related issues. An argument for the former is that simple verdicts cost less time to read and are easier to comprehend, compared with the complex policy-related issues. Another argument for

the former is that targeting politicians' statements hold them accountable and deter them from lying. Some studies find evidence that fact-checking reduces lying behaviors of politicians (e.g., Nyhan and Reifler (2015); Lim (2018)). This paper provides a theoretical ground for the deterrence argument and shows that the public can be better off under lie detection in spite of the ignorance of details in policy-related issues.

Related Literature: The present paper is mostly related to the literature of strategic communication with lie detection. Balbuzanov (2019) and Dziuda and Salas (2018) analyze a cheap-talk model akin to the setup in Crawford and Sobel (1982), with the addition that the sender's lie will be detected with an exogenous probability.[3] The information structures of the signal the receiver observes are identical between these two. Balbuzanov (2019) shows that given an intermediate probability of lie detection and a sufficiently small bias, fully revealing equilibria exist. Dziuda and Salas (2018) show that certain refinement criteria lead to a unique equilibrium where the moderate and high types stay honest and the low types lie to imitate the high types.

Our findings in the optimal equilibrium echo findings from Dziuda and Salas (2018) that moderate types do not exaggerate their types to avoid being mistaken as the low-type liars. The key difference between the present paper and previous literature is that we model lie detection as a decision of the receiver, where the probability of lie detection can be chosen conditional on the sender's claim. This allows an analysis of tensions between the sender's incentive of lying and the receiver's incentive of inspection. Dziuda and Salas (2018) also has a different signal structure of lie detection compared to the present model. In their model, if the sender tells the truth the receiver observes the message the sender chose, and this message appears consistent to the receiver. If the sender who observes the state θ sends a lying message with a literal meaning of "the state is θ_l", this message appears inconsistent to the receiver with probability $p > 0$, and appears consistent to the receiver with probability $1 - p$. Under this signal structure, the receiver learns that the messages are lies for certain through inconsistent messages, while consistent messages might still indicate pooling between truth-tellers and liars. Dziuda and Salas (2018) showed the existence of equilibria that have the following structures. In equilibrium, the moderate and the high types tell the truth, while the low types lie claiming to be the high types. Specifically, there are two cutoffs t and l in the state space $(0 < t < l < 1)$ such that if the state is

[3] Ederer and Min (2022) consider a model of Bayesian persuasion in which the receiver detects the sender's lies with positive probability. Similar to Balbuzanov (2019) and Dziuda and Salas (2018), the signal structure of lie-detecting technology is exogenously given.

above t, the sender tells the truth in a consistent way; otherwise, the sender randomly sends lying messages in a consistent way and claims that the state is above l. Hence, in equilibrium, the receiver detects the state if it is in (t, l). Similar to our model, a fear of implicit lying cost is an important key to dissuade the moderate and the high types from sending lying messages. Under the prescribed sender's message strategy, from the receiver's viewpoint, messages appeared in an inconsistent way are the evidence of the state is below t. On the other hand, even if messages correspond to the high types appeared in a consistent way, the receiver is still confused between the corresponding truth teller and some liars. Then, she chooses an action as if the state arises on the boundary l. Since they assumed that the sender wants the receiver to choose an action that is as high as possible, this implies that lying messages may induce an unfavorable action with a positive probability. Therefore, the sender of type $\theta > t$ tells the truth to remove the possibility that the receiver observes an inconsistent message.

Jehiel (2021) analyzes an interesting multi-round cheap-talk environment where lie can be spotted from the inconsistent messages of a forgetful liar who cannot remember the content of the lie he has told. We can interpret the observation of inconsistencies by the receiver as a lie detection technology. Similar to our equilibrium construction, the fear of being inconsistent causes the sender to be more careful in his pronouncements. Jehiel (2021) shows that as the state space becomes finer, lies will be detected for sure and, thus, all equilibria in pure strategies approximate the fully revealing equilibrium. Hence, there is no inconsistency in equilibrium. In contrast, we show that if the communication is one shot and the sender's lie is detected through the receiver's strategic lie detection, lying and lie detection must occur on the equilibrium path to improve the receiver's welfare.

Levkun (2021) study a sender-receiver game with a strategic fact-checker. In contrast to our model, the state and decision are binary, and a strategic third party (fact-checker) checks whether the sender's message is true. Further, the fact-checker can commit to the fact-checking policy in advance. It is shown that if the cost of lie detection is small, the optimal lie-detecting policy for the fact-checker is full lie detection; otherwise, no lie detection is optimal. In contrast, we show that with multiple states and no commitment on the inspection policy, partial lie detection is always optimal when the cost of lie detection is not too high.

While other researchers have studied similar cheap-talk models with a partially informed receiver, such as Ball and Gao (2019), Chen (2009), Chen (2012), Ishida and Shimizu (2016), Ishida and Shimizu (2019), Moreno de Barreda (2013), Miyahara and Sadakane (2020), Lai (2014), and Rantakari (2016), the present study differs in its information structure. Similar to Balbuzanov (2019) and Dziuda

and Salas (2018), these prior studies assume that the receiver obtains additional signals except for the sender's message. However, these signal are either exogenous or they do not provide certifiable information that detects the sender's lie.

Ishida and Shimizu (2019) study a cheap-talk model in which the receiver is partially informed about the state. They assume that the receiver has limited knowledge of the signal structure. Specifically, the receiver faces higher-order uncertainty, and the signal that she observes is noisy. In this setting, the sender's message can help the receiver remove the noise of the private signal to some extent (i.e., *confirmation effect*). In contrast, in the present study, the sender's message strategy induce a noise, and lie detection helps the receiver remove the noise added by the lying types.

The present study is also closely related to the literature on strategic communication games with lying costs (e.g., Kartik et al. (2007); Kartik (2009); Nguyen and Tan (2019); Guo and Shmaya (2020)).[4] The key difference is that the sender's lying cost function is exogenously given in the literature. One interpretation of this assumption is that the sender faces the psychological costs of lying.[5] In contrast, in our model, the costs that the sender bears from lying are endogenously determined in equilibrium. The lying costs come from the receiver's response depending on the misunderstanding, rather than the sender's psychological costs.[6] Consequently, while lying behaviors arise in equilibrium in both these models and ours, the natures and interpretations of lies are quite different. In our model, lies serve as disguises to confuse the receiver. Liars try to mimic the types they claim to be, and the receiver cannot tell them apart without inspection. In their models, lies serve as inflated languages. The sender tells a lie to avoid being mistaken as a worse type, and a strategic receiver does not confuse a liar with the type he claims to be.

A signaling equilibrium with similar structures exists in literature. Feltovich et al. (2002) study

[4] An alternative interpretation of the models in Kartik et al. (2007), alongside other related works (e.g. Ottaviani and Squintani (2006); Chen (2011)) is that a proportion of receivers naively believes sender's message. The coexistence of strategic and naive receivers imposes an endogenous cost for the sender to overly exaggerate the state since the naive receivers will take it at face value, which is not preferred by a sender whose bias is not too large. In the equilibria of their models, lies are chosen by the sender to balance the induced beliefs of two groups of receivers who interpret messages differently. In our model, lies are chosen to mimic the corresponding truthful senders and confuse the receiver.

[5] Examples of experimental evidence include Gneezy (2005), Lundquist et al. (2009), Fischbacher and Föllmi-Heusi (2013), López-Pérez and Spiegelman (2013), Abeler et al. (2014), Nyhan and Reifler (2015), Abeler et al. (2019).

[6] Austen-Smith and Banks (2000) and Kartik (2007) analyze a model of cheap talk and costly signaling. The signaling instrument takes the form of "burning money," and its cost does not vary with the sender's type. The authors' focus is on how the additional option of "burning money" can enlarge the set of CS equilibria.

a sender-receiver game, assuming that the sender can send a costly signal and the receiver observes additional noisy exogenous information on the state. They have shown that there is an equilibrium in which medium types signal to separate themselves from low types, but high types choose to not signal, that is, low types and high types pool together. The high types can save costs by not signaling (counter signaling) and relying on the additional information to stochastically separate them from low types. Moreover, the counter signaling itself is a signal that separates high types from medium types. This equilibrium shares a similar structure with ours. However, the incentives behind the equilibrium are different to ours because the sender cannot send a costly signal and the receiver has to make a costly effort in our model.

The remainder of this paper is organized as follows. Section 2 presents the model. Section 3 derives necessary and sufficient conditions for the existence of welfare-improving lie-detecting equilibrium. Section 4 characterizes the optimal lie-detecting equilibrium. Section 5 compares lie detection with state verification. Section 6 discusses the optimality of inspected vague messages under more general distributions. Section 7 concludes. The proofs are relegated to the Appendices.

2 The Model

There are two players, a receiver and a sender. The receiver has to make a decision, but only the sender has the relevant information. The sender privately observes the state of the world, θ, which is distributed over a normalized state space $\Theta \equiv [0,1]$ with a twice continuously differentiable c.d.f. F, with the associated density function f such that $f(\theta) > 0$ for all $\theta \in [0,1]$ (full support). θ is also referred to as the sender's type. For example, θ might represent the quality of the advertised product or the severity of crimes committed by a suspect.

Message: We adopt the same definition of message and lying in Sobel (2020). The sender sends a message $m \in \mathcal{M}$ to the receiver, where $\mathcal{M}$ is associated with the Borel $\sigma-$algebra of the state space Θ: for every Borel $\Theta_0 \subseteq \Theta$, there exists a message m_{Θ_0} that implies $\theta \in \Theta_0$.[7] A message sent by the sender is interpreted as a statement regarding his type. To simplify the notation, as long as there is no risk of confusion, we denote m_{Θ_0} by $m = \Theta_0$. To provide a few examples, a message $m = [0.3, 0.4]$

[7] $\mathcal{M}$ is defined as a compact metrizable space. The Borel $\sigma-$algebra on Θ, denoted by $\mathbb{B}(\Theta)$, has a cardinality of continuum. Therefore, for example, we can take $\mathbb{R}$ as $\mathcal{M}$. When we provide the formal discussions on the measurability of the sender's message strategy, the probability distribution of messages, and the conditional distribution on the state space given messages, we adopt $(\mathcal{M}, \mathbb{B}(\mathcal{M}))$ as the measurable space.

is interpreted as the following statement: "my type lies somewhere in between 0.3 and 0.4"; a message $m = \{0.5\} \cup \{0.7\}$ is interpreted as "my type is either 0.5 or 0.7"; a message $m = \Theta$ can be interpreted as to remain silent because it essentially means "Anything is possible."

Costly inspection: The receiver, after observing m, chooses whether or not to inspect the message with a cost $c > 0$. An inspection reveals the truthfulness of the statement. Formally, if an inspection takes place, the receiver will receive a binary signal such that for every Borel $m \subseteq \Theta$,

$$s(m, \theta) = \begin{cases} t & \text{if } \theta \in m \\ l & \text{otherwise.} \end{cases} \tag{1}$$

If the receiver chooses not to inspect, she receives an uninformative signal $s(m, \theta) = u$. The signal t indicates the sender's message is inspected and confirmed to be truthful; l indicates the sender's message is inspected and confirmed to be a lie; u indicates the sender's message is uninspected.

Action: After observing both the message m and the inspection signal s, the receiver chooses a payoff relevant action $x \in [0, 1]$.

Preference: The receiver has a quadratic loss function $u_r(x, \theta) = -(x - \theta)^2 - cI$, where $I = 1$ if an inspection took place, $I = 0$ otherwise. The sender has a von Neumann-Morgenstern utility $u_s(x)$ which is strictly increasing in x. In other words, there is no common interest between the receiver and sender. The receiver wants to take an action that matches the true state, while the sender always prefers an action that is as high as possible, independent of the true state.

Strategy profile: A strategy profile (q, P, X) consists of of three measurable maps; a message strategy $q : \Theta \to \mathcal{M}$, where $q(\theta)$ is the message sent by type θ [8]; an inspection rule $P : \mathcal{M} \to [0, 1]$, where $P(m)$ is the probability for the receiver to inspect message m; and an action rule $X : \mathcal{M} \times \{t, l, u\} \to [0, 1]$, where $X(m, s)$ is the action taken following message m and inspection signal $s \in \{t, l, u\}$.

Given a message strategy q, let $\mathcal{M}_q = q(\Theta)$ be the set of all on-path messages.[9] For any on-path

[8] For expositional clarity, we confine attention to pure message strategies in this paper, i.e. each type of the sender θ sends a message $q(\theta)$ with probability 1. In Appendix C, we show that the results in this paper can be generalized to allow mixed message strategy.

[9] Throughout this paper, we follow the convention and refer to $g(X)$ as $\{y : \exists x \in X \text{ such that } y \in g(x)\}$ for any function or correspondence g and set X within the domain of g.

message $m \in \mathcal{M}_q$, let

$$\Theta_q^t(m) = \{\theta \in \Theta : q(\theta) = m \text{ and } \theta \in m\} \tag{2}$$

$$\Theta_q^l(m) = \{\theta \in \Theta : q(\theta) = m \text{ and } \theta \notin m\} \tag{3}$$

$$\Theta_q^u(m) = \Theta_q^t(m) \cup \Theta_q^l(m) \tag{4}$$

be the sets of truthful senders, lying senders, and senders of m, respectively. The receiver cannot commit to an inspection rule and/or an action rule. They have to be sequentially rational based on a Bayesian updated belief. Given q and $M \subseteq \mathcal{M}_q$, let $\Theta_q^s(M) = \bigcup_{m \in M} \Theta_q^s(m)$. Then, we uniquely obtain the regular conditional expectation $E[\theta|\theta \in \Theta_q^s(M)]$ for $s \in \{t, l, u\}$, where $\mu(\Theta')E[\theta|\theta \in \Theta'] = \int_{\Theta'} \theta dF(\theta)$; $E[\theta|\theta \in \Theta']$ denotes the conditional expected type given a set of type $\Theta' \subseteq \Theta$, and $\mu(\Theta') = \int_{\Theta'} dF(\theta)$ denotes the probability of Θ'.[10] Let $w_q(m) = E[\mathbf{1}_{\Theta_q^l(m)}|\Theta_q^u(m)]$ be the conditional probability of the sender being a liar given that he sends m.[11] If $\mu(\Theta_q^u(m)) > 0$, $w_q(m) = \mu(\Theta_q^l(m))/\mu(\Theta_q^u(m))$. Let $Var(\Theta') = E[(\theta - E[\theta|\Theta'])^2|\Theta'] = E[\theta^2|\Theta'] - \{E[\theta|\Theta'\}^2$ be the conditional variance given Θ'. Then, $Var(\Theta') = \frac{\int_{\Theta'}(\theta - E[\Theta'])^2 dF(\theta)}{\mu(\Theta')}$ if $\mu(\Theta') > 0$.

Given a message strategy q, we define the message distribution $H_q(m)$ such that for any subset of equilibrium messages $M \subseteq \mathcal{M}_q$,[12]

$$\int_M dH_q(m) = \int_{\Theta_q^u(M)} dF(\theta).$$

Sequentially rational action: Since the receiver's utility is quadratic, her optimal action is equal to the conditional expectation of the sender's type given the posterior belief; therefore an action strategy

[10] Let $(\Theta, \mathbb{B}(\Theta), \mu)$ be the probability space on which F is defined. Let $\mathcal{F}$ be the σ-algebra generated by $\Theta_q^s(\cdot)$ for $s \in \{t, l, u\}$ and $M \in \mathbb{B}(\mathcal{M})$. A conditional expectation of θ given $\mathcal{F}$, denoted $E[\theta|\mathcal{F}]$, is any $\mathcal{F}$-measurable function which satisfies $\int_{\tilde{\Theta}} E[\theta|\mathcal{F}]d\mu = \int_{\tilde{\Theta}} \theta d\mu$ for $\tilde{\Theta} \in \mathcal{F}$. For $s \in \{t, l, u\}$ and any Borel subset of on-path messages $M \subseteq \mathcal{M}_q$ such that $\mu(\Theta_q^s(M)) > 0$, we obtain $\mu(\Theta_q^s(M))E[\theta|\theta \in \Theta_q^s(M)] = \int_{\Theta_q^s(M)} \theta d\mu(\theta)$. Then, $E[\theta|\theta \in \Theta_q^s(m)]$ is almost surely unique.

[11] Given the probability space $(\Theta, \mathbb{B}(\Theta), \mu)$ on which F is defined, the sender's measurable message strategy q is a $(\mathcal{M}, \mathbb{B}(\mathcal{M}))$-valued random variable. Then, the regular conditional probability, $Pr(\theta \in \tilde{\Theta}|q = m) = E[\mathbf{1}_{\tilde{\Theta}}|\theta \in \Theta_q^u(m)]$ for $\tilde{\Theta} \in \mathbb{B}(\Theta)$, is defined as a function $\nu : \mathcal{M} \times \mathbb{B}(\Theta) \to [0, 1]$ such that (i) for almost every $m \in \mathcal{M}$, $\nu(m, \cdot)$ is a probability measure on $(\Theta, \mathbb{B}(\Theta))$; (ii) for all $\tilde{\Theta} \in \mathbb{B}(\Theta)$, $\nu(\cdot, \tilde{\Theta})$ is $\mathbb{B}(\mathcal{M})$-measurable, and (iii) for all $\tilde{\Theta} \in \mathbb{B}(\Theta)$ and $M \in \mathbb{B}(\mathcal{M})$, $\mu(\tilde{\Theta} \cap q^{-1}(M)) = \int_M \nu(m, \tilde{\Theta})\mu(q^{-1}(dm))$. Therefore, a more precise condition for $w_q(m)$ is that $w_q(m) = \nu(m, \Theta_q^l(m)) = E[\mathbf{1}_{\Theta_q^l(m)}|\Theta_q^u(m)]$.

[12] The message distribution H_q is uniquely introduced by the probability measure $\mathcal{H}_q$ on $(\mathcal{M}, \mathbb{B}(\mathcal{M}))$ such that $\mathcal{H}_q(M) = \mu(q^{-1}(M))$ for every $M \in \mathbb{B}(\mathcal{M})$.

X is **sequentially rational** given q if for any $m \in \mathcal{M}_q$ and $s \in \{t, l, u\}$,

$$X(m, s) = E[\Theta_q^s(m)], \tag{5}$$

where $E[\Theta_q^s(m)]$ denotes $E[\theta|\theta \in \Theta_q^s(m)]$.

After observing the on-path message m and inspection signal s, the receiver chooses an action to match the conditional expected type of senders who send m and lead to inspection signal s given the message rule q. Instead of blindly taking a message at its face value, a Bayesian, sequentially rational receiver updates her belief given the set of equilibrium senders who would pass/fail an inspection, and reacts optimally. When there is no inspection, the receiver remains aware of the possibility of lying and chooses an action that matches the weighted average type of the equilibrium truth-tellers and liars.

Information value of inspection: Given a message strategy q and a sequentially rational action strategy X, the receiver's expected continuation payoff under the quadratic utility function **if she inspects an on-path message** $m \in \mathcal{M}_q$ is:

$$-w_q(m)Var(\Theta_q^l(m)) - (1 - w_q(m))Var(\Theta_q^t(m)).$$

Recall that $w_q(m)$ is the conditional probability of the sender of m being a liar. Upon inspection, the receivers expected loss from action imprecision for a message m is the weighted average conditional variance of equilibrium truth-tellers and liars of m.

The receiver's expected continuation payoff if she does not inspect m is:

$$-Var(\Theta_q^u(m)),$$

which is the variance of the sender's type conditional on him sending m. Since $\Theta_q^u(m) = \Theta_q^t(m) \cup \Theta_q^l(m)$, the law of total variance implies that

$$\begin{aligned} Var(\Theta_q^u(m)) =& w_q(m)Var(\Theta_q^l(m)) + (1 - w_q(m))Var(\Theta_q^t(m)) + w_q(m)(1 - w_q(m))(E[\Theta_q^l(m)]^2 + E[\Theta_q^t(m)]^2) \\ & - 2w_q(m)(1 - w_q(m))E[\Theta_q^l(m)]E[\Theta_q^t(m)] \\ =& w_q(m)Var(\Theta_q^l(m)) + (1 - w_q(m))Var(\Theta_q^t(m)) + w_q(m)(1 - w_q(m))(E[\Theta_q^l(m)] - E[\Theta_q^t(m)])^2. \end{aligned}$$

Therefore, the information value of inspecting m is the reduction in conditional variance from the binary signal:

$$\begin{aligned} V_q(m) =& Var(\Theta_q^u(m)) - w_q(m)Var(\Theta_q^l(m)) - (1 - w_q(m))Var(\Theta_q^t(m)) \\ =& w_q(m)(1 - w_q(m))(E[\Theta_q^l(m)] - E[\Theta_q^t(m)])^2. \end{aligned} \tag{6}$$

An inspection allows the receiver to make a better inference on the sender's type and chooses more precise action accordingly. If there is a large difference between the expected type of truth-tellers and liars who send m, the value of differentiating these two groups is large. Besides, an inspection is more informative when the liar to truth-teller ratio is less extreme. If the sender of m is very likely to be on one side, not much information is revealed from an inspection. An inspection strategy P is **sequentially rational** given q if for any $m \in \mathcal{M}_q$,

$$P(m) \in \begin{cases} \{0\} & \text{if } c > V_q(m) \\ [0,1] & \text{if } c = V_q(m) \\ \{1\} & \text{if } c < V_q(m). \end{cases} \tag{7}$$

In other words, the receiver will inspect only if the information value of inspection is no less than the cost of inspection.

Sender's optimality: Given inspection strategy P and action strategy X, type θ sender's expected utility from sending a message m is:

$$EU_{X,P}(m|\theta) = \begin{cases} P(m)u_s(X(m,t)) + (1-P(m))u_s(X(m,u)) & \text{if } \theta \in m \\ P(m)u_s(X(m,l)) + (1-P(m))u_s(X(m,u)) & \text{if } \theta \notin m. \end{cases} \tag{8}$$

A message strategy q is **optimal** given P and X if for any $\theta \in \Theta$ and $m' \in \mathcal{M}_q$,[13]

$$EU_{X,P}(q(\theta)|\theta) \geq EU_{X,P}(m'|\theta). \tag{9}$$

3 Equilibrium analysis

This section defines a Perfect Bayesian equilibrium and establishes the necessary and sufficient conditions for the existence of an equilibrium where inspections take place with positive probability.

Definition 1 *A strategy profile $\sigma = (q, P, X)$ is a Perfect Bayesian equilibrium if P and X are sequentially rational given q, and q is optimal given P and X.*

Given an equilibrium σ, the receiver's ex-ante expected payoff is:

[13] Incentive constraints over off-path messages are omitted because sequential rationality put no restriction on the inspections and actions following those messages. Therefore, we can without loss of generality let $X(m', s) = 0$ for any off-path message m', and sender will have no incentive to deviate to those messages.

$$EU_r(\sigma) = -\int_{\mathcal{M}_q} (1-P(m))(X(m,u)-\theta)^2 + P(m)[w_q(m)(X(m,l)-\theta)^2 + (1-w_q(m))(X(m,t)-\theta)^2 + c]dH_q(m). \quad (10)$$

Define

$$G_\sigma(x) = \int_{\mathcal{M}_q} (1-P(m))\mathbf{1}(X(m,u)\leq x) + P(m)[w_q(m)\mathbf{1}(X(m,l)\leq x) + (1-w_q(m))\mathbf{1}(X(m,t)\leq x)]dH_q(m) \quad (11)$$

be the distribution of induced actions under σ, and

$$p_\sigma = \int_{\mathcal{M}_q} P(m)dH_q(m), \quad (12)$$

be the ex-ante probability that the sender is inspected under σ. Sequential rationality of the action rule X implies that

$$\begin{aligned} EU_r(\sigma) &= \int_{\mathcal{M}_q} (1-P(m))(X(m,u)^2 - E[\theta^2|\Theta_q^u(m)]) \\ &\quad + P(m)[w_q(m)(X(m,l)^2 - E[\theta^2|\Theta_q^l(m)]) + (1-w_q(m))(X(m,t)^2 - E[\theta^2|\Theta_q^t(m)]) - c]dH_q(m) \\ &= \int_{\mathcal{M}_q} (1-P(m))X(m,u)^2 + P(m)[w_q(m)X(m,l)^2 + (1-w_q(m))X(m,t)^2] \\ &\quad - cP(m) - E[\theta^2|\Theta_q^u(m)]dH_q(m) \\ &= \int_0^1 x^2 dG_\sigma(x) - cp_\sigma - E[\theta^2], \end{aligned} \quad (13)$$

where $E[\theta^2] \equiv \int_\Theta \theta^2 dF(\theta)$. The sender's ex-ante expected payoff is:

$$EU_s(\sigma) = \int_0^1 u_s(x)dG_\sigma(x). \quad (14)$$

We refer to the pair (G_σ, p_σ) as the **induced outcome distribution** of the equilibrium σ. We say two equilibria σ and σ' are **distribution equivalent** if they have the same induced outcome distribution. Since (G_σ, p_σ) uniquely determine payoffs in an equilibrium, two distribution equivalent equilibria induce the same expected payoffs for the receiver and every type of the sender.

Since the receiver cannot commit to a sub-optimal action rule, the expected value of induced actions must equal the expected value of the state. In fact, the distribution of induced actions G is a mean-preserving contraction of the prior distribution F. A more dispersed G implies a more precise match between the induced actions and the states, and thus a higher expected payoff for the receiver.

Proposition 1 *For any equilibrium $\sigma = (q, P, X)$ there exists a distribution equivalent equilibrium $\hat{\sigma} = (\hat{q}, \hat{P}, \hat{X})$ such that for any $m \in \mathcal{M}_{\hat{q}}$:*
(i) $\hat{X}(m,t) \geq \hat{X}(m,l)$, and
(ii) $m = \Theta^t_{\hat{q}}(m)$.

Condition (i) of Proposition 1 provides a natural interpretation of an equilibrium: liars attempt to pool with truth-tellers in the hope of inducing higher actions. A liar who send m induces $\hat{X}(m,u)$ when the message is uninspected, which is higher than $\hat{X}(m,l)$ only if condition (i) is satisfied.[14] Condition (ii) comes from the fact that condensing the statement of a message to include only equilibrium truth-tellers is the most effective equilibrium construction in order to minimize the sender's incentive of deviation. To illustrate this idea, suppose a message m is sent by type 0.3 (the equilibrium liar) and type 0.8 (the equilibrium truth-teller). Now consider two equilibria where other things being equal, except $m = \{0.7, 0.8\}$ is the first equilibrium (so that condition (ii) does not hold) and $m = \{0.8\}$ in the second equilibrium (so that condition (ii) does not hold). If type 0.7 deviated from his equilibrium message to m, he would have passed an inspection and induce 0.8 in the first equilibrium but failed an inspection and induce 0.3 in the second equilibrium. Therefore, if he has no incentive to deviate to m in the first equilibrium, he will have no incentive to deviate to m in the second equilibrium. Proposition 1 is useful in analyzing the set of implementable outcome because an outcome distribution is implementable if and only if it can be induced by an equilibrium that satisfies the above properties.[15] Unless otherwise stated, any equilibrium discussed henceforth satisfies conditions (i) and (ii) of Proposition 1.

Fix an equilibrium $\sigma = (q, P, X)$ and let $\mathcal{M}^0_q = \{m \in \mathcal{M}_q : P(m) = 0\}$ be the set of on-path uninspected messages. The sender's optimality requires that all messages in $\mathcal{M}^0_q$ must induce the same action. Otherwise, senders who induce a lower uninspected action will deviate to a higher one. Therefore, we can without loss assume that there is at most one such message, m^0_q, and all senders

[14] In a model where sender can make a truthful claim and trick the lie detector to identify him as a liar (e.g., by acting nervous or intentionally failing a test), then condition (i) must hold in any equilibrium for any inspected message m, for otherwise equilibrium truth-tellers who act normally and get $X(m,t)$ will deviate to act nervously and get $X(m,l)$.

[15] Note however that oftentimes an implementable outcome distribution can also be induced by other equilibria. For example, if there exists an on-path message m' which is never inspected, and Θ' is the set of senders of m', an equilibrium that satisfies condition (ii) requires the statement m' to be a subset of Θ'. However, it would still be an equilibrium if senders of m' simply " remain silent," i.e., $m' = \Theta$. By definition, it means every type becomes a truth-teller of m', but it has no effect on the sender's incentive because being truthful or lying makes no difference to the outcome when m' is never inspected.

of that message are truthful, i.e. $m_q^0 = \Theta_q^u(m) = \Theta_q^t(m) \equiv \Theta_q^0$, where Θ_q^0 is the set of types who are never inspected in equilibrium. Sequential rationality of X requires $X(m_q^0, u) = E[\Theta_q^0]$. Let $\mathcal{M}_q^+ = \{m \in \mathcal{M}_q : P(m) > 0\}$ be the set of messages that are inspected with positive probability. $\mathcal{M}_q^+$ is simply referred to as the **set of inspected messages**. For $\theta \in \Theta$, we say θ is **truthful** if $\theta \in \Theta_q^t(\mathcal{M}_q^+) = \{\theta : P(q(\theta)) > 0 \text{ and } \theta \in q(\theta)\}$; θ is **lying** if $\theta \in \Theta_q^l(\mathcal{M}_q^+) = \{\theta : P(q(\theta)) > 0 \text{ and } \theta \notin q(\theta)\}$; θ is **uninspected** if $\theta \in \Theta_q^0 = \{\theta : P(q(\theta)) = 0\}$.

It is clear that under lie-detecting technology, a babbling equilibrium (where every type pools into a single message and the receiver never inspect) always exists. Moreover, the babbling outcome will be the unique outcome if the receiver never inspects in an equilibrium. An equilibrium outcome can be different from the babbling outcome only if some messages are inspected in equilibrium.

We say σ is an **equilibrium with inspection** if $p_\sigma > 0$, i.e, some on-path messages are inspected with positive probability.

3.1 An Example of Equilibrium with Inspection

Assume that the sender's type is uniformly distributed and the inspection cost is $c = \frac{25}{288}$. Consider the following message strategy profile with two equilibrium messages: $m_I = [\frac{1}{2}, 1]$ sent by the set of types $[0, \frac{1}{4}] \cup [\frac{1}{2}, 1]$, and $m_U = (\frac{1}{4}, \frac{1}{2})$ sent by the set of types $(\frac{1}{4}, \frac{1}{2})$. Therefore, the set $\Theta_t \equiv [\frac{1}{2}, 1]$ is the set of truth-tellers of m_I while the set $\Theta_l \equiv [0, \frac{1}{4}]$ is the set of liars of m_I. A sequentially rational action strategy profile is $X(m_I, t) = \frac{3}{4}$, $X(m_I, l) = \frac{1}{8}$, $X(m_I, u) = \frac{13}{24}$, and $X(m_U, u) = X(m_U, t) = X(m_U, l) = \frac{3}{8}$. Using equation (6) it is straightforward to verify that the value of inspecting m_I is $\frac{25}{288}$ and the value of inspecting m_U is 0. Therefore, the receiver is indifferent between inspecting m_I or not, and strictly prefers not to inspect m_U. Now consider an inspection strategy profile: $P(m_I) = \frac{u_s(X(m_I,u)) - u_s(X(m_U,u))}{u_s(X(m_I,u)) - u_s(X(m_I,l))}$ and $P(m) = 0$ for any $m \neq m_I$. Note that $P(m_I) \in (0, 1)$ as $X(m_I, u) > X(m_U, u) > X(m_I, l)$. Given $P(m_I)$, the sender is indifferent between sending m_U (getting $X(m_U, u)$ with certainty) and being a liar at m_I (getting $X(m_I, u)$ with probability $1 - P(m_I)$ or $X(m_I, l)$ with probability $P(m_I)$); therefore, the types in $[0, \frac{1}{4}]$ and $(\frac{1}{4}, \frac{1}{2})$ have no incentive to deviate, while the types in $[\frac{1}{2}, 1]$ receive a strictly higher payoff and, therefore, they have no incentive to deviate either. Hence, the above strategy profile is an equilibrium with inspection.

Generally speaking, an equilibrium with inspection is supported by two kinds of equilibrium messages: suspicious messages and an innocent message. The suspicious messages are sent by some low-type liars and high-type truth-tellers in a proportion that makes the receiver indifferent between

inspecting the messages or not, while the innocent message is sent by some medium types where the receiver finds it unworthy to inspect. The receiver stochastically inspected the suspicious messages in a way that makes the sender indifferent between lying in the suspicious messages and sending the innocent message.

The following subsection establishes the necessary and sufficient conditions for the existence of an equilibrium with inspection.

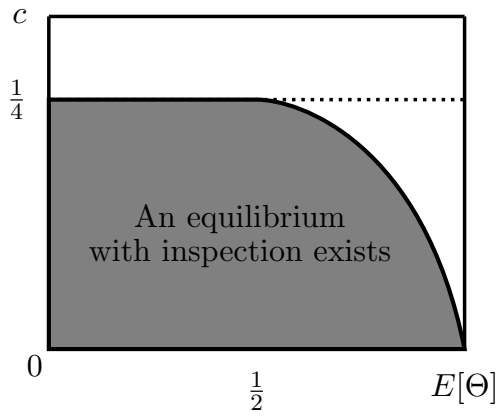
3.2 Existence of Equilibrium with Inspection

Assumption 1 $c < \frac{1}{4}$ *and* $E[\Theta] \equiv \int_0^1 \theta dF(\theta) < \frac{1}{2} + \sqrt{\frac{1}{4} - c}$.

Proposition 2 *There exists an equilibrium with inspection if and only if Assumption 1 is satisfied. Moreover, the receiver gets a higher ex-ante expected payoff in that equilibrium compared with the babbling outcome.*

The credibility of inspections relies on the existence of both liars and truth-tellers. Upon receiving a message, if the receiver's expectation on the sender's type is extreme (either too high or too low), the information value of an inspection is low because the sender is either very likely to be truth-telling or very likely to be lying, and thus, any inspection is non-credible. Now consider an uninspected message m and a randomly inspected message m'. In order to incentivize the liars who send m' to take the risk of being caught, the receiver's expectation on the sender's type upon receiving m' must be higher than the expectation upon receiving m, so that if liars of m' get away with the lie, they receive a higher payoff than those who send m. Since the receiver is Bayesian, her expectation upon receiving m' must be higher than the prior expectation. Therefore, if the prior expectation is too optimistic, her expectation upon receiving m' will also be too optimistic for the inspection to be credible. It is worth noting that the condition is not symmetric. Even if prior expectation on the sender's type is pessimistic, it is possible to design an equilibrium with pessimistic belief for the uninspected message and moderate beliefs for the inspected messages so that liars of the inspected messages are incentivized and the inspections are credible. Therefore, the lie-detecting technology is useful when the prior expectation is moderate or pessimistic, but not when it is optimistic, i.e., $E[\Theta]$ is too high given c. Figure 1 depicts the region of parameter values in which an equilibrium with a positive probability of inspection exists. The threshold of prior expectation such that an inspection is credible is decreasing in the cost of inspection, meaning that when the cost is smaller, an inspection is credible for a larger range of optimistic beliefs.

Figure 1: Parameter values that allow the existence of an equilibrium with inspection



When the cost of inspection is small, inspections are credible even if conditional expectations given the inspected statements are optimistic and the information values of inspection are small. Inspection can therefore facilitate information transmission. As cost goes to 0, the lie-detecting technology is useful for almost any prior distribution.

4 Receiver-optimal Equilibrium

This section defines the receiver-optimal equilibrium and establishes its properties.

Definition 2 *An equilibrium σ is receiver-optimal if for any equilibrium σ', $EU_r(\sigma) \geq EU_r(\sigma')$.*

Unless otherwise specified, we will simply refer a receiver-optimal equilibrium to as an **optimal equilibrium**. An optimal equilibrium induces the highest expected payoff to the receiver among all equilibria. We focus on analyzing the best equilibrium for the receiver because oftentimes the receiver's welfare reflects the public interest, for instance, consumers and voters who have to make decisions under incomplete information. The optimal equilibrium indicates an upper bound to the welfare of the public under lie-detecting technology. Besides, the optimal equilibrium minimizes a weighted average objective of the inference error and the inspection cost. Therefore, it can be interpreted as the most efficient way of combating disinformation using lie-detecting technology. On the other hand, the sender's welfare is sensitive to his risk attitude. It is worth noting that if the sender is risk neutral, he will get the same ex-ante payoff in any equilibrium, because the mean of the induced action

distribution must equal the prior expectation of the state. In such a case, the outcome induced by an optimal equilibrium is also Pareto-efficient.

Now we derive some properties of an optimal equilibrium. We say a property holds **almost everywhere** for a set of messages M if it holds for a subset of messages $M' \subseteq M$ such that $H_q(M') = H_q(M)$.

Proposition 3 *In any optimal equilibrium σ, $V_q(m) = c$ almost everywhere for $m \in \mathcal{M}_q^+$.*

Recall that $w_q(m)$ is the conditional probability of the sender being a liar given a message m.

Proposition 4 *In any optimal equilibrium σ, $w_q(m) \leq 0.5$ almost everywhere for $m \in \mathcal{M}_q^+$.*

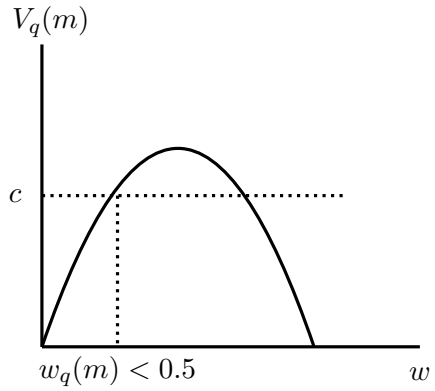
The value of lie-detecting technology to the receiver is composed of two parts: direct information value and indirect deterrence effect. Proposition 3 says that direct information value of inspection is offset by the cost of inspection in any optimal equilibrium, and the net benefit of inspection comes from its effect on the sender's incentive: some types of sender refrain from making a higher claim because of the possible lie detection. As a result, some information is transmitted through the messages in the sense that expectations of the sender's type upon receiving different messages are different; therefore, the receiver is able to make a better inference on the sender's type even when an inspection does not take place ex-post. Proposition 4 says that for any inspected message in the optimal equilibrium, liars are a minority compared with truth-tellers. It is because any inspected message in an equilibrium requires a moderate liar to truth-teller ratios so that information values are high enough for credible inspections. Such ratios can be achieved by either a minority or a majority of liars. Compared with an equilibrium with a majority of liars, an equilibrium with a minority of liars means that the expected type of the sender of inspected messages are higher. That creates larger differences between conditional expectations between the inspected messages and the uninspected message, which means more information is transmitted through messages under an equilibrium with a minority of liars. Proposition 3 and Proposition 4 together imply that an optimal equilibrium minimizes the proportion of liars subject to the constraint that $V(m) \geq c$, as depicted in Figure 2.

4.1 Optimal equilibrium under uniform distribution: Characterization

In this subsection, we concentrate on the well-known uniform distribution and provide further characterization of the optimal equilibrium.[16]

[16] In a previous version of this paper, we show that part of the results provided in this subsection can be extended to more general distributions. For details, see Proposition 5 and Proposition 6 of Tam(2019).

Figure 2: The optimal proportion of liars



Assumption 2 $F(\theta) = \theta$.

Now we derive an upper bound of the receiver's payoff in any equilibrium.

Proposition 5 *For any equilibrium σ under Assumption 2 and $c \leq \frac{1}{4}$, $EU_r(\sigma) \leq -\frac{\alpha^3}{12} - c(1-\alpha)$, where $\alpha = 2\sqrt{c}$.*

In any equilibrium, the sender's type space can be divided into two sets: the uninspected types (Θ_q^0) and the (randomly) inspected types (Θ_q^+). Since there can only be a single action induced by the set of uninspected types, the best-case scenario for the receiver is that Θ_q^0 is an interval with some length α. In such cases, the expected loss is $\frac{\alpha^3}{12}$. For the inspected types Θ_q^+, the best-case scenario for the receiver is that those types are perfectly revealed upon inspection. In such cases, the expected cost of inspection is $c(1-\alpha)$. The total expected loss is minimized when the length of the set of uninspected types is $\alpha = 2\sqrt{c}$.

In the followings, we construct a strategy profile called **decreasing mimicking strategy** and show that under the uniform distribution, it is an equilibrium that achieves the upper bound payoff in Proposition 5 and, thus, an optimal equilibrium.

For $d \in [2\sqrt{c}, 1]$, define

$$w^-(d) = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{c}{d^2}}, \tag{15}$$

which is the smaller root of the equation $w(1-w)d^2 = c$. Recall that the value of inspecting a message m is $w_q(m)(1-w_q(m))(E[\Theta_q^l(m)] - E[\Theta_q^t(m)])^2$. Therefore, provided that d is the distance between

the conditional expected type of truth-tellers and liars in a message m, $w^-(d)$ will be the minimum proportion of liars such that the information value of inspecting m is no less than c. This minimum proportion is decreasing in d, meaning that the credibility of inspection can be sustained for a smaller proportion of liars when the distance between the two conditional expectations is larger. Note that $2\sqrt{c}$ is the minimum required distance such that an inspection can be made credible, and $w^-(2\sqrt{c}) = \frac{1}{2}$. Proposition 3 and Proposition 4 imply for any inspected message $m \in \mathcal{M}_q^+$ in an optimal equilibrium,

$$w_q(m) = w^-(X(m,t) - X(m,l)), \tag{16}$$

Thus, for any inspected message in an optimal equilibrium, the proportion of liars is uniquely determined by the distance between expected types of truth-tellers and liars. For $x_l \in [0, 1-2\sqrt{c}]$ and $x_t \in [x_l + 2\sqrt{c}, 1]$, define

$$X_u^*(x_t, x_l) = w^-(x_t - x_l)x_l + (1 - w^-(x_t - x_l))x_t, \tag{17}$$

which is the expected type of senders of a message m where x_t is the expected type of truth-tellers, x_l is the expected type of liars, and the proportion of liars is minimized subject to the receiver's incentive constraint of inspection. Since the receiver is sequentially rational, for any $m \in \mathcal{M}_q^+$ in an optimal equilibrium, we obtain

$$X(m,u) = X_u^*(X(m,t), X(m,l)). \tag{18}$$

Therefore, when the inspection does not take place ex-post, the induced action is uniquely determined by the expected type of truth-tellers and liars.

Now we define the decreasing mimicking strategy. A pair of cutoffs and matching bijection ($\underline{\theta}_d$, $\overline{\theta}_d$, and $\phi_d : [\overline{\theta}_d, 1] \to [0, \underline{\theta}_d]$) is defined as a solution of the following system of differential equation and boundary conditions:

$$\dot{\phi}_d(\theta) = -\frac{w^-(\theta - \phi_d(\theta))}{1 - w^-(\theta - \phi_d(\theta))} \tag{19}$$

$$\phi_d(1) = 0 \tag{20}$$

$$\phi_d(\overline{\theta}_d) = \underline{\theta}_d = \overline{\theta}_d - 2\sqrt{c}, \tag{21}$$

where $\phi_d(.)$ represents a decreasing matching bijection from the truthful interval to the lying interval which specifies the lying pattern in the decreasing mimicking strategy. That is, $\phi(\theta)$ represents a liar who mimics θ. The pair of boundaries $(\underline{\theta}_d, \overline{\theta}_d)$ is pinned down by the initial condition (20), differential

equation (19), and the terminal condition that the distance between the two boundaries is $2\sqrt{c}$. Note that for $c < \frac{1}{4}$, $w^-(1) \in (0, \frac{1}{2})$ is well defined, so is ϕ_d. We have $\overline{\theta}_d - \phi_d(\overline{\theta}_d) = 2\sqrt{c} < 1 = 1 - \phi_d(1)$; therefore, $\overline{\theta}_d < 1$ is well defined and $\underline{\theta}_d = \phi_d(\overline{\theta}_d) > 0$ because ϕ_d is strictly decreasing.

Define the **decreasing mimicking strategy** σ_d which is characterized by $(\underline{\theta}_d, \overline{\theta}_d, \phi_d)$ defined in conditions (19) - (21) as follows:

(i) **Intermediate types - Uninspected vague claim**: There is an uninspected message $m_q^0 = [\underline{\theta}_d, \overline{\theta}_d]$ sent by $\theta \in [\underline{\theta}_d, \overline{\theta}_d]$ and $P(m_q^0) = 0$.

(ii) **High types - Randomly inspected, precise claims**: There is a continuum of randomly inspected messages $\mathcal{M}_q^+ = \{m = \{\theta\} : \theta \in (\overline{\theta}_d, 1]\}$, each $m \in \mathcal{M}_q^+$ sent by the truthful type $\theta = m$ and $P(m) \in (0, 1)$.

(iii) **Low types - Mimicking the high types**: Each $m \in \mathcal{M}_q^+$ is sent by a liar $\phi_d(m)$.

The action strategy X is determined by sequential rationality. For $m \in \mathcal{M}_q^+$,

$$\begin{aligned} X(m, t) &= m \\ X(m, l) &= \phi_d(m) \\ X(m, u) &= X_u^*(m, \phi_d(m)), \end{aligned} \tag{22}$$

and

$$X(m_q^0, u) = E[\underline{\theta}_d, \overline{\theta}_d]. \tag{23}$$

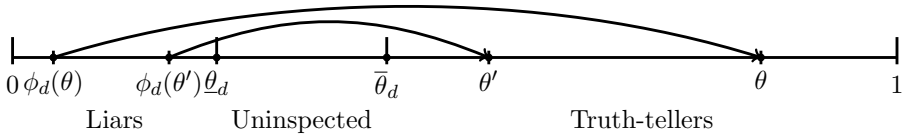
The inspection strategy P for $m \in \mathcal{M}_q^+$ is determined by the incentive compatibility conditions of the liars:

$$P(m) = \frac{u_s(X(m, u)) - u_s(X(m_q^0, u))}{u_s(X(m, u)) - u_s(X(m, l))}. \tag{24}$$

Figure 3 depicts the structure of the decreasing mimicking strategy. Under σ_d, each truthful type θ makes the precise claim "My type is θ," and each of such claim is mimicked by exactly one type of liar $\phi_d(\theta)$, where $\phi_d(.)$ is decreasing; therefore, worse liars tell bigger lies. Upon receiving each of these messages, the receiver is indifferent between inspecting and not inspecting. This is true because for any $m \in (\overline{\theta}_d, 1]$,

$$\begin{aligned} \frac{w_q(m)}{1 - w_q(m)} &= \lim_{\epsilon \to 0} \frac{\mu([\phi_d(m), \phi_d(m - \epsilon)])}{\mu([m - \epsilon, m])} \\ &= -\dot{\phi}_d(m) = \frac{w^-(m - \phi_d(m))}{1 - w^-(m - \phi_d(m))}, \end{aligned}$$

Figure 3: The structure of decreasing mimicking equilibrium σ_d.



where the first equality holds because of the continuously decreasing message strategy, and the third equality holds by (19).[17] Therefore, $w_q(m) = w^-(m - \phi_d(m))$ and thus $V_q(m) = c$. The slope of the matching bijection $\phi_d(.)$ is chosen so that for each $m \in (\overline{\theta}_d, 1]$, the value of inspecting m equals c given the probability densities of the truth-tellers and liars at the points m and $\phi_d(m)$. The inspection probability is chosen so that liars are indifferent between telling such lies and making the uninspected claim. Requiring a precise statement for high claims helps in making more precise decisions upon inspection. Random inspections of those claims are justified because each of them is made by a low type and a high type. A vague moderate claim pools the moderate types which are not distant enough to be worth inspecting.

Proposition 6 *If Assumption 1 and Assumption 2 are satisfied, the decreasing mimicking strategy σ_d is an optimal equilibrium, that is, $EU_r(\sigma_d) = -\frac{\alpha^3}{12} - c(1-\alpha)$, where $\alpha = 2\sqrt{c}$. Besides, $X(m,u)$ is increasing in m for $m \in (\overline{\theta}_d, 1]$.*

The optimal equilibrium σ_d has a three-interval structure such that when the state is above the cutoff $\overline{\theta}_d$, the sender is truthful; when the state is below the cutoff $\underline{\theta}_d$, sender lies and claims that the state is somewhere above $\overline{\theta}_d$, such claims are inspected with positive probabilities; when the state is intermediate, sender makes the claim in which the receiver does not inspect. Such structure induces

[17] The sender's strategy for low and high types is a measurable function $q_d : [0, \underline{\theta}_d) \cup (\overline{\theta}_d, 1] \to (\overline{\theta}_d, 1]$ such that $q_d^{-1}(m) = \{\phi_d(m)\} \cup \{m\}$. Since q_d is $((\overline{\theta}_d, 1], \mathbb{B}((\overline{\theta}_d, 1]))$-valued random valuable, for any $[m, m'] \subset (\overline{\theta}_d, 1)$ and $\Theta' \in \mathbb{B}([0, \underline{\theta}_d) \cup (\overline{\theta}_d, 1])$, we must have

$$\mu(\theta \in \Theta' | q(\theta) \in [m, m']) = \frac{\mu(\theta \in \Theta' \cap q_d^{-1}([m, m']))}{\mu(\theta \in q_d^{-1}([m, m']))}.$$

Note that $\phi_d([m, m'])$ and $[m, m']$ belong to $\mathbb{B}([0, \underline{\theta}_d) \cup (\overline{\theta}_d, 1])$; $\phi_d([m, m']) \cap q_d^{-1}([m, m']) = \phi_d([m, m']) = [\phi_d(m'), \phi_d(m)]$; and $[m, m'] \cap q_d^{-1}([m, m']) = [m, m']$. Therefore,

$$\frac{\mu(\theta \in [\phi_d(m'), \phi_d(m)] | q(\theta) \in [m, m'])}{\mu(\theta \in [m, m'] | q(\theta) \in [m, m'])} = \frac{\phi_d(m) - \phi_d(m')}{m' - m}.$$

Hence, $\frac{w_q(m)}{1 - w_q(m)} = \lim_{m' \to m} \frac{\phi_d(m) - \phi_d(m')}{m' - m} = -\dot{\phi}_d(m)$.

disperse inferences upon inspection, which benefit the receiver the most. Under the optimal equilibrium, low type senders are incentivized to lie in order to justify inspections of the truthful statements made by high type senders. Such inspections prevent moderate type senders from exaggerating their types in the fear of getting caught lying and being perceived as low types. A higher message m induces a higher action $X(m,u)$ when m is uninspected. It means that the sender's message conveys useful information to the receiver even when the message is not inspected ex-post. The types who send a lower message has no incentive to deviate to a higher one because in the case of an inspection, a liar will be punished by a lower induced action $X(m,l)$ if he sends a higher message m, and the inspection probability $P(m)$ could be higher for a higher message m.[18]

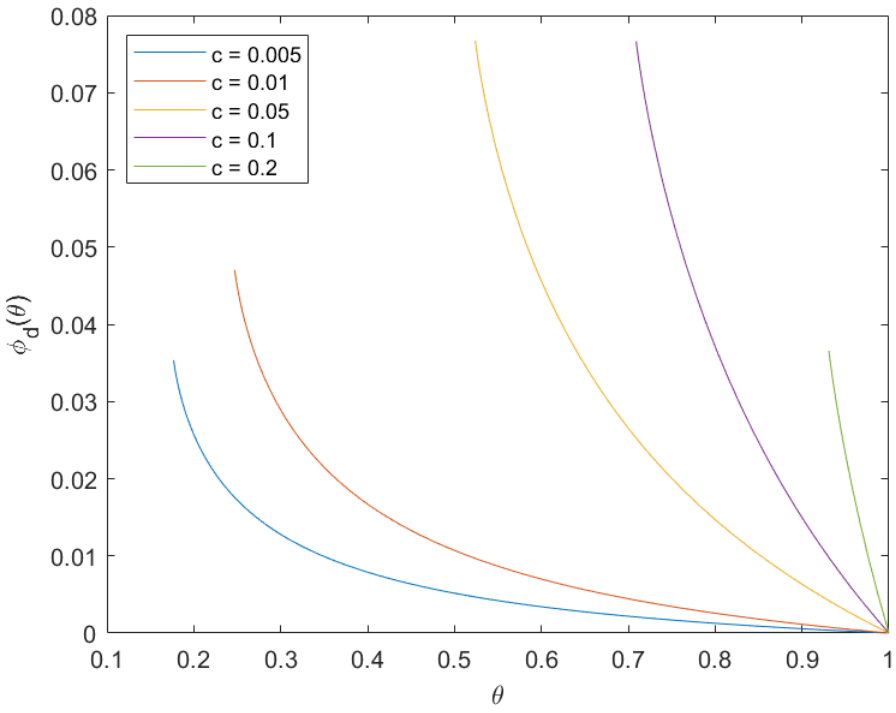
Dziuda and Salas (2018) also find an equilibrium with a three-interval structure in their model with exogenous lie detection. There are two main differences between the information structures of their equilibrium and the optimal equilibrium found in the present paper. First, the intermediate types are separating in their equilibrium, whereas the intermediate types pool into a single message in our equilibrium due to the cost saving consideration specific to the present model with endogenous and costly inspection. Second, when a lie is detected in their equilibrium, the receiver has the same posterior belief on the sender's type regardless of the claim made by the liar. In our equilibrium, the sender's type is separating upon any equilibrium inspection. It is achievable under endogenous lie detection because the probabilities of inspection can be used as a degree of freedom to manipulate the sender's payoff. Given the appropriate inspection probabilities, the sender is indifferent between each equilibrium message even though different lies induce different actions upon inspection.

It is worth mentioning that the decreasing mimicking equilibrium is not the unique optimal equilibrium under the uniform distribution. For instance, an increasing mimicking equilibrium where the matching bijection is increasing can also be optimal. We focus on the decreasing mimicking equilibrium due to its ease of computation. We are able to solve differential equation (19) with a known boundary condition $\phi_d(1) = 0$, and derive the numerical results presented in Figure 4 and Figure 5. Propositions 5 and 6 imply that any optimal equilibrium has a three-interval structure with a one-to-one mimicking mapping from the low interval to the high interval, and an uninspected middle interval with a length equals $2\sqrt{c}$.

Figure 4 and Figure 5 depict the matching bijection $\phi_d(.)$ and the cutoffs $(\overline{\theta}_d, \underline{\theta}_d)$ in the decreasing

[18] Generally speaking, $P(m)$ is not necessarily increasing in m. Its shape also depends on the sender's utility function, according to the equilibrium condition (24).

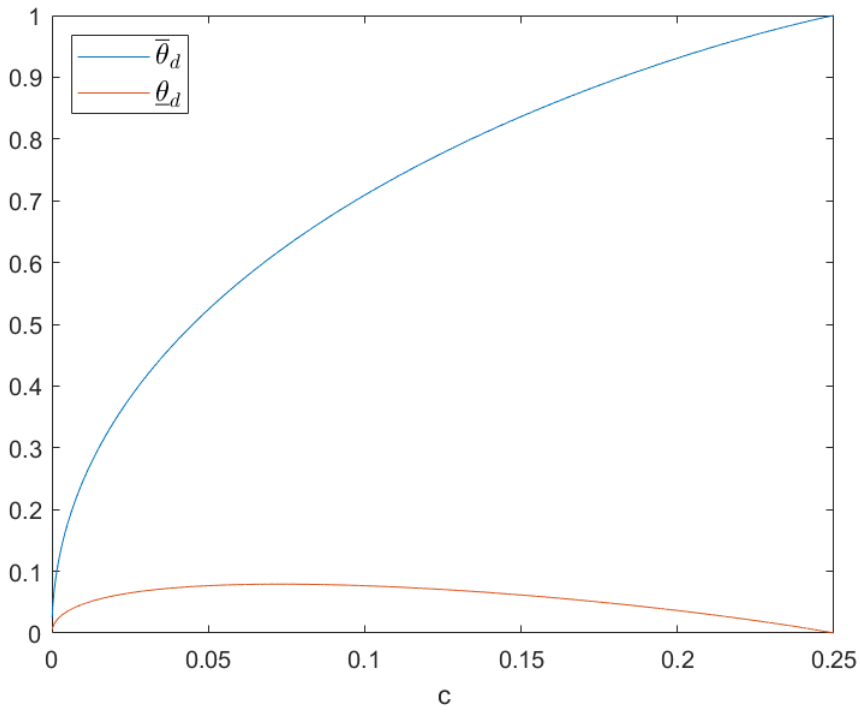
Figure 4: The matching bijection ϕ_d in the decreasing mimicking strategy (Please print in color).



mimicking strategy under different levels of inspection costs. The slope of the decreasing matching bijection $\phi_d(.)$ is steeper for a larger inspection cost, since a larger mass of liars is needed to match with each unit mass of truth-tellers in order to maintain the credibility of inspection. The upper cutoff $\overline{\theta}_d$ is increasing in c while the lower cutoff $\underline{\theta}_d$ is non-monotonic in c. The inspection cost affects the two cutoffs through two forces. First, the optimal length of the uninspected interval is $2\sqrt{c}$; therefore, the larger the c, the shorter the intervals of truth-tellers and liars. Second, the larger the c, the larger the liar to truth-teller ratio is. For small c, the second effect dominates the first effect; therefore, $\underline{\theta}_d$ is increasing in c; For large c, the first effect dominates the second effect, so $\underline{\theta}_d$ is decreasing in c. It is worth noting that $\underline{\theta}_d$ is below 0.1 for any inspection cost. The proportion of lying senders necessary to sustain an optimal lie-detecting equilibrium is relatively small. As c goes to 0, both $\overline{\theta}_d$ and $\underline{\theta}_d$ converge to 0, the sender is almost always truth-telling. The optimal lie-detecting equilibrium approximates the full information outcome.

For $c < 0.25$, the receiver's expected payoff under the decreasing mimicking strategy is $-\frac{(2\sqrt{c})^3}{12} - c(1-2\sqrt{c})$, where the first term is the expected loss from the uninspected interval and the second term is the expected cost of inspecting the upper and lower intervals. The receiver's payoff is higher for a

Figure 5: The cutoffs $\overline{\theta}_d$ and $\underline{\theta}_d$ in the decreasing mimicking strategy (Please print in color).



lower inspection cost and it converges to 0 as c goes to zero.

4.2 Robustness of the Optimal Equilibrium under Message Dependent Inspection Costs

In the above analysis we assume that the receiver's inspection cost is independent of the message sent by the sender. One might argue that realistically the inspection cost should be higher for vague messages because they are harder to prove or disprove. For instance, consider the inspection technology where the receiver can choose some states and verify whether the chosen states are the true state one by one. The larger the set of states she has to go through, the higher the inspection cost. The decreasing mimicking equilibrium would still be an optimal equilibrium under such an inspection technology because every inspected message in that equilibrium is precise and, thus, costs the least to inspect.

5 Discussion: State verification and lie detection

In this section, we compare state-verifying technology and lie-detecting technology, in particular, the receiver's welfare under the two technologies. Instead of revealing a binary signal as in (1), consider that the true state is revealed upon inspection; therefore, by paying cost c to inspect the message m, the receiver obtains a precise signal

$$s(m, \theta) = \theta. \tag{25}$$

If the receiver chooses not to inspect, she receives an uninformative signal $s(m, \theta) = u$. Under state-verifying technology, the sequentially rational action rule for the receiver is

$$X(m, \theta) = \theta; X(m, u) = E[\Theta_q^u(m)], \tag{26}$$

where $\Theta_q^u(m)$ is the set of senders who send m, and value of verifying m is the conditional variance of the sender's type:

$$V_q(m) = Var(\Theta_q^u(m)), \tag{27}$$

and the sequentially rational inspection rule for the receiver is

$$P(m) \in \begin{cases} \{0\} & \text{if } c > V_q(m) \\ [0,1] & \text{if } c = V_q(m) \\ \{1\} & \text{if } c < V_q(m). \end{cases} \tag{28}$$

Type θ sender's expected utility from sending a message m is

$$EU_{X,P}(m|\theta) = P(m)u_s(\theta) + (1 - P(m))u_s(X(m, u)), \tag{29}$$

and the sender's optimality implies that for any $\theta \in \Theta$ and on-path message $m' \in \mathcal{M}_q$,

$$P(q(\theta))u_s(\theta) + (1 - P(q(\theta)))u_s(X(q(\theta), u) \geq P(m')u_s(\theta) + (1 - P(m'))u_s(X(m', u), \tag{30}$$

where $q(\theta)$ is the message sent by θ under the equilibrium. We will show that there are only two kinds of equilibria under costly state verification.

Uninformative equilibrium: $P(m) = 0$ and $X(m, u) = E[\Theta]$ for any $m \in \mathcal{M}_q$, and

State-verifying equilibrium: $P(m) = 1$ and $X(m, \theta) = \theta$ for any $m \in \mathcal{M}_q$.

Proposition 7 *Under the costly state-verifying technology, if $c > Var(\Theta)$, only uninformative equilibrium exists; if $c < Var(\Theta)$, only state-verifying equilibrium exists.*

The ability to reveal the state precisely upon an inspection completely eliminates any incentive for the sender to transmit information. Gain from state-verifying technology comes solely from the direct information value. It is contrary to the lie-detecting technology, which benefits the receiver by manipulating the sender's incentive to transmit information. Such manipulation is possible because the nature of lie detection creates a strategic uncertainty to the receiver: even if she spots a lie, she does not reveal the true type of the liar and has to decide the action based on equilibrium inference. This could benefit the receiver in an ex-ante sense because the sender might be deterred from deviation in the fear of being mistaken as a worse type than what he actually is, and such a deterrence effect facilitates informative communication. However, if the receiver reveals the true state from an inspection, this deterrence will not be credible, and there will be no reason for the sender to stay honest. As a result, revealing more information from the inspection eliminates voluntary information transmission from the sender. Although a similar effect of the receiver's more accurate information on the sender's information transmission has been found in previous studies, e.g. Chen (2012), Ishida and Shimizu (2016), Ispano (2016), and Dziuda and Salas (2018) , the present paper provides its welfare implication on two costly inspection technologies the explain the underlying mechanics of the effect in these specific environments.

With Proposition 7, the receiver's ex-ante payoff under the costly state-verifying technology is

$$EU_r^v = -\min\{Var(\Theta), c\}. \tag{31}$$

The following Proposition shows that learning more from the inspection reduces the receiver's payoff.

Proposition 8 *Let σ^* be the optimal equilibrium under the lie-detecting technology. Then under any inspection cost, $EU_r(\sigma^*) \geq EU_r^v$. Furthermore, if $c < Var(\Theta)$, then $EU_r(\sigma^*) > EU_r^v$.*

The optimal equilibrium under lie-detecting technology outperforms the state-verifying equilibrium due to the possibility of discriminative inspection. The receiver is able to inspect the more important claims while leaving the less important claim uninpsected for cost-saving measure. This result provides a theoretical foundation for the emphasis on a sender's integrity, instead of objective information. By neglecting further information about the truth (other than the information that determines whether the sender is lying), the receiver is able to impose a credible threat that whoever being caught lying will

be perceived poorly, regardless of the sender's true type. Therefore, even though there is no common interest between sender and receiver, some types of sender refrain from making higher claims in the fear of being perceived as a worse type than they actually are.

6 Discussion: Inspected vague messages in an optimal equilibrium

One implication of Proposition 6 is that under the uniform distribution, every message that is inspected with positive probability in the optimal equilibrium is "precise," which means that inspected messages are never vague. It suggests that vague messages are not necessary in achieving the best outcome for the receiver under the lie-detecting environment[19]. It is true since all inspected messages are precise, and the uninspected message can also be precise because it is never inspected anyway; thus, the content of the message does not matter.

One might wonder if sending precise messages is a general property of an optimal lie-detecting equilibrium under any distribution, and the answer is no. For simple exposition, consider the following discrete distribution:

Example: A three-type distribution, $\theta \in \{0, 0.8, 1\}$.

Prior distribution: $f \begin{pmatrix} 0 \\ 0.8 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.1 \\ 0.1 \end{pmatrix}$; Inspection cost: $c = 0.2025$.

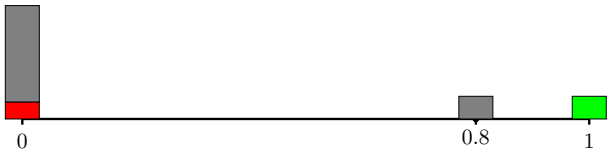
Note that under this inspection cost, it is credible for the receiver to inspect a message with truth-tellers from type 1 and liars from type 0, given an appropriate weight of liars. However, it is non-credible for the receiver to inspect a message with truth-tellers from type 0.8 and liars from type 0, regardless of the weight. First consider the optimal equilibrium in which the sender can only send precise messages.

Optimal equilibrium with precise messages: type 1 sends the truthful message $m^+ = 1$, type 0.8 sends another arbitrary message m^0, and type 0 randomizes between m^+ (with probability 0.0491) and m^0 (with probability 0.9509) that makes the inspection of m^+ credible. In this equilibrium, m^+ will be inspected with probability determined according to (24) and m^0 will be uninspected. Figure 6 depicts the messaging structure of the above equilibrium. The receiver's payoff is -0.085. While it is

[19] In Appendix B we show that with the restriction of deterministic inspection, inspected vague messages can be optimal under uniform distribution

beneficial to separate type 0.8 from the uninspected message m^0, there is no credible way to achieve that using precise messages. Now suppose the sender is allowed to send vague messages.

Figure 6: Optimal equilibrium under a three-type distribution when messages are restricted to be precise (Please print in color).



Optimal equilibrium with a vague message: type 1 and type 0.8 pool into a vague message $m_v^+ = \{0.8, 1\}$, so that both types are truth-tellers of the message m_v^+; type 0 randomizes between m_v^+ (with probability $\frac{1}{4}$) and another arbitrary message m^0 (with probability $\frac{3}{4}$). In this equilibrium, m_v^+ will be inspected with probability 1 and m^0 will be uninspected. Figure 7 depicts the messaging structure of the above equilibrium. The receiver's payoff $= -0.083 > -0.085$.

By pooling type 0.8 and type 1 into a vague message m_v^+, the average truth-telling type of m_v^+ is 0.9, which makes it credible to inspect m_v^+ when paired with type 0 liars. By allowing vague messages, it is possible to separate type 0.8 from m^0.

In a previous version of this paper, we show that the decreasing mimicking equilibrium with precise messages is receiver-optimal when the distribution is negatively skewed or symmetrical (see Proposition 6 and Remark 1 of Tam(2019) for details). The above example suggests that when the distribution is positively skewed, the optimal equilibrium might consist of vague inspected messages. An insight from this observation is that when the market is flooded with low quality products, having a quality standard that covers a wider range of high quality types might be beneficial to the consumers, because it provides sufficient incentive for the consumers to verify the standard and separate a wider range of

Figure 7: Optimal equilibrium under a three-type distribution when vague messages are allowed (Please print in color).

high quality products from the low quality products.

7 Conclusion

We establish a framework that allows analyses on the strategic interaction between cheap talk and lie detection, and characterize the receiver's optimal equilibrium. The results suggest that optimal lie detection works as a credible deterrence tool. Low types are induced to lie so that inspections are justified, which deter higher types from lying. Under certain conditions, such optimal equilibrium can be achieved by allowing the sender to choose among a vague moderate claim and a continuum of precise high claims. This provides a direction for efficient allocation of resources in combating misinformation in various aspects such as politics and product advertising.

Several potential extensions are worth mentioning. As one of the early attempts in the literature to study endogenous lying and costly lie detection, we restrict attention to the setting of single round communication and lie detection. In some applications, the sender and the receiver can conduct multiple rounds of communication and lie detection, before a final decision is made by the receiver. For instance, the police can ask the suspect multiple questions and conduct lie detection for each claim made by the suspect. Dziuda and Salas (2018) show that the receiver prefers to commit to a single round communication when the probability of lie detection is exogenously high, because anticipating the second chance of communication makes the sender more likely to lie. It might appear that this effect is strengthened when lie detection is costly as the receiver has to pay the cost of inspection in each round. A formal analysis is required for such an argument. Another potential extension is to allow a certain degree of common interest between the sender and the receiver, such as biased sender as in the CS model. If the cost of lie detection is small, at least one deterministic lie-detecting equilibrium characterized in an online appendix exists, because it always exists as long as the sender is upwardly biased. However, it is not clear how having a sender with a smaller bias would change the receiver's optimal equilibrium, especially its strategic effect on the stochastic lie detection is elusive. On one hand, sender with smaller bias is willing to reveal more precise information, as suggested by the standard cheap-talk model. On the other hand, when bias is small, there is no way to induce the sender to tell big lies. This hinders the formation of credible inspection. Without inspection, the sender might be tempted to tell small lies, which impede informative communication. The analysis of these opposing effects may present interesting avenues for future research.

References

Abeler, J., Becker, A., and A. Falk (2014), "Representative evidence on lying costs," *Journal of Public Economics*, 113, 96–104.

Abeler, J., Nosenzo, D., and C. Raymond (2019), "Preferences for truth-telling," *Econometrica*, 87(4), 1115–1153.

Argenziano, R., S. Severinov, and F. Squintani (2016), "Strategic information acquisition and transmission," *American Economic Journal: Microeconomics*, 8(3), 119–155.

Austen-Smith, D. (1994), "Strategic transmission of costly information," *Econometrica*, 62, 955–963.

Austen-Smith, D. and J. S. Banks (2000), "Cheap talk and burned money," *Journal of Economic Theory*, 91(1), 1–16.

Balbuzanov, I. (2019), "Lies and Consequences: The Effect of Lie Detection on Communication Outcomes," *International Journal of Game Theory*, 48(4), 1203–1240.

Ball, I. and Gao, X. (2019), "Checking Cheap Talk," Working paper.

Chen, Y. (2009), "Communication with two-sided asymmetric information," Working paper, available at SSRN 1344818.

Chen, Y. (2011), "Perturbed communication games with honest senders and naive receivers," *Journal of Economic Theory*, 20(1), 146(2), 401–424.

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

Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E. and K. B. McDermott (2008), "The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses," *Cerebral cortex*, 19(7), 1557–1566.

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

Dziuda, W. and C. Salas (2018), "Communication with detectable deceit," Working paper, Available at SSRN 3234695.

Ederer, F. and W. Min (2022). "Bayesian Persuasion with Lie Detection," No. w30065. National Bureau of Economic Research.

Feltovich, N., Harbaugh, R., and To, T. (2002). "Too cool for school? Signalling and countersignalling," *RAND Journal of Economics*, 630–649.

Fischbacher, U. and F. Föllmi-Heusi (2013), "Lies in disguise: an experimental study on cheating," *Journal of the European Economic Association*, 11(3), 525–547.

Gneezy, U. (2005), "Deception: The role of consequences," *American Economic Review*, 95(1), 384–394.

Guo, Y. and E. Shmaya (2020), "Costly miscalibration," *Theoretical Economics*, forthcoming.

Hartwig, M. and C. F. Bond Jr (2014), "Lie detection from multiple cues: A meta-analysis," *Applied Cognitive Psychology*, 28(5), 661–676.

Hartwig, M., Granhag, P. A., Stromwall, L. A. and O. Kronkvist (2006), "Strategic use of evidence during police interviews: When training to detect deception works," *Law and human behavior*, 30(5), 603–619.

Intergovernmental Panel on Climate Change (2018), "Special Report: Global Warming of 1.5°C," (https://www.ipcc.ch/sr15/).

Ishida, J. and T. Shimizu (2016), "Cheap talk with an informed receiver," *Economic Theory Bulletin*, 4(1), 61–72.

Ishida, J. and T. Shimizu (2019), "Cheap talk when the receiver has uncertain information sources," *Economic Theory*, 68(2). 303–334.

Ispano, A. (2016), "Persuasion and receiver's news," *Economics Letters*, 141, 60-63.

Jehiel, P. (2021), "Communication with Forgetful Liars." *Theoretical Economics*, 16(2), 605–638.

Kartik, N. (2007), "A note on cheap talk and burned money," *Journal of Economic Theory*, 136(1), 749–758.

Kartik, N. (2009), "Strategic communication with lying costs," *The Review of Economic Studies*, 76(4), 1359–1395.

Kartik, N., Ottaviani, M. and F. Squintani (2007), "Credulity, lies, and costly talk," *Journal of Economic theory*, 134(1), 93–116.

López-Pérez, R. and E. Spiegelman (2013), "Why do people tell the truth? Experimental evidence for pure lie aversion," *Experimental Economics*, 16(3), 233–247.

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

Levkun, A. (2021), "Communication with Strategic Fact-checking," Working paper.

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

Lundquist, T., Ellingsen, T., Gribbe, E., and M. Johannesson (2009), "The aversion to lying. Journal of Economic Behavior & Organization," 70(1-2), 81–92.

Moreno de Barreda, I. (2013), "Cheap talk with two-sided private information," Working Paper.

Miyahara, Y. and H. Sadakane (2020), "Communication enhancement through information acquisition by uninformed player," KIER Discussion Paper No. 1050.

Nguyen, A. and T. Y. Tan (2019), "Bayesian persuasion with costly messages," Working Paper, available at SSRN 3298275.

Nyhan, B. and J. Reifler (2015), "The effect of fact-checking on elites: A field experiment on US state legislators," *American Journal of Political Science*, 59(3), 628–640.

Ottaviani, M. and F. Squintani (2006), "Naive audience and communication bias," *International Journal of Game Theory*, 35(1), 129–150.

Pei, D. (2015), "Communication with endogenous information acquisition," *Journal of Economic Theory*, 160, 132–149.

Porter, M. and F. Ten Brinke (2008), "Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions," *Psychological science*, 19(5), 508–514.

Rantakari, H. (2016), "Soliciting advice: Active versus passive principals," *The Journal of Law, Economics, and Organization*, 32(4), 719–761.

Sadakane, H. (2020), "Cheap talk and Fact-checking," Working Paper.

Sobel, J. (2020), "Lying and Deception in Games," *Journal of Political Economy*, 128(3), 907–947.

Tam, T. Y. C. (2019), "Lying and Lie-detecting," Working Paper.

Vrij, A., Granhag, P. A., Mann, S. and S. Leal (2011), "Outsmarting the liars: Toward a cognitive lie detection approach," *Current Directions in Psychological Science*, 20(1), 28–32.

Appendix A

This Appendix provides proofs of Propositions 1 – 8.

Proof of Proposition 1:

Fix an equilibrium $\sigma = (q, P, X)$. Let $M_1 = \{m \in \mathcal{M}_q : X(m,l) > X(m,t)\}$ be the set of on-path messages such that the induced action of liars is higher than the induced action of truth-tellers. Define a modified equilibrium $\hat{\sigma} = (\hat{q}, \hat{P}, \hat{X})$ such that for each $m \in \mathcal{M}_q/M_1$, the set of senders remains unchanged but they now send the transformed message $T(m) = \Theta_q^t(m)$. For each $m \in M_1$, the set of senders remains unchanged but they now send the transformed message $T(m) = \Theta_q^l(m)$. For inspection probabilities, let $\hat{P}(T(m)) = P(m)$ for each $m \in \mathcal{M}_q$.

The set of on-path messages of the modified equilibrium $\hat{\sigma}$ is $\mathcal{M}_{\hat{q}} = T(\mathcal{M}_q)$. The sequentially rational actions for $\hat{\sigma}$ are $\hat{X}(T(m), s) = X(m, s)$ for $m \in \mathcal{M}_q/M_1$ and $s = t, l, u$; $\hat{X}(T(m), t) = X(m, l)$, $\hat{X}(T(m), l) = X(m, t)$ and $\hat{X}(T(m), u) = X(m, u)$ for $m \in M_1$. It is straightforward that for all $m \in \mathcal{M}_q$ $\hat{X}(T(m), t) \geq \hat{X}(T(m), l)$ and $T(m) = \Theta_q^t(T(m))$. Therefore, condition (i) and (ii) are satisfied in the modified equilibrium $\hat{\sigma}$. Furthermore, since the induced actions remain unchanged for every type of sender, so $\hat{\sigma}$ and σ are distribution equivalent.

To see that $\hat{\sigma}$ is an equilibrium, note that for $m \in \mathcal{M}_q/M_1$, $w_q(m) = w_{\hat{q}}(T(m))$, and for $m \in M_1$, $w_q(m) = 1 - w_{\hat{q}}(T(m))$. Therefore, for any $m \in \mathcal{M}_q$, $V_q(m) = w_q(m)(1-w_q(m))(X(m,t)-X(m,l))^2 = w_{\hat{q}}(T(m))(1-w_{\hat{q}}(T(m)))(\hat{X}(T(m),t)-\hat{X}(T(m),l))^2 = V_{\hat{q}}(T(m))$, thus the receiver's optimal inspection condition (7) remains satisfied in $\hat{\sigma}$. To check the sender's optiamlity condition (9), note that the equilibrium payoff of each type of sender remains unchanged, i.e. $EU_{\hat{X},\hat{P}}(\hat{q}(\theta)|\theta) = EU_{X,P}(q(\theta)|\theta)$. By the definition of the modified set of message, any type θ would be identified as a liar of any on-path message other than its equilibrium message, i.e. $\theta \notin m'$ for any $m' \in T(\mathcal{M}_q)/\hat{q}(\theta)$. This combined with the fact that $\hat{X}(T(m), t) \geq \hat{X}(T(m), l)$ implies $EU_{\hat{X},\hat{P}}(T(m')|\theta) \leq EU_{X,P}(m'|\theta)$ for any $m' \in \mathcal{M}_q$. Therefore, $EU_{\hat{X},\hat{P}}(\hat{q}(\theta)|\theta) = EU_{X,P}(q(\theta)|\theta) \geq EU_{X,P}(m'|\theta) \geq EU_{\hat{X},\hat{P}}(T(m')|\theta)$ for any $\theta \in \Theta$ and $m' \in \mathcal{M}_q$, where the first inequality holds by the optimality of the original equilibrium σ, thus (9) is satisfied in the modified equilibrium. Therefore, we conclude that $\hat{\sigma}$ is an equilibrium.

Q.E.D.

The following Lemma establishes the necessary and sufficient conditions of an equilibrium, which are useful in proving the subsequent propositions.

Lemma 1 *Let q and X be a pair of message and action strategies that satisfy (i) and (ii) of Proposition*

1, X satisfies the receiver's sequential rationality (5) given q, and let m_q^0 be the (potentially non-existent) uninspected message. Then there exists an inspection strategy P on the set of inspected messages M_q^+ such that (q, P, X) is an equilibrium if and only if for any $m, m' \in \mathcal{M}_q^+$:

(a) $X(m,l) \leq X(m_q^0, u) < X(m', u)$ if m_q^0 exists; $X(m,l) < X(m', u)$ otherwise;

(b) $w_q(m)(1-w_q(m))(X(m,t)-X(m,l))^2 \begin{cases} = c & \text{if } X(m,l) < X(m_q^0, u) \\ \geq c & \text{if } X(m,l) = X(m_q^0, u) \end{cases}$; if m_q^0 does not exist, replace $X(m_q^0, u)$ with $\sup_{m' \in \mathcal{M}_q^+} X(m', l)$.

In particular, $P(m) = \frac{u_s(X(m,u))-u_s(x_q^0)}{u_s(X(m,u))-u_s(X(m,l))}$, where $x_q^0 = X(m_q^0, u)$ if m_q^0 exists;

$x_q^0 \in [\sup_{m' \in \mathcal{M}_q^+} X(m', l), \inf_{m'' \in \mathcal{M}_q^+} X(m'', u)]$ if m_q^0 does not exist and $V_q(m) = c$ for all $m \in \mathcal{M}_q^+$;

$x_q^0 = \max_{m' \in \mathcal{M}_q^+} X(m', l)$ otherwise.

Proof of Lemma 1:

Given (i) of Proposition 1, we have $X(m,t) \geq X(m,l)$, and for any $m \in \mathcal{M}_q^+$, it must be the case that $X(m,t) > X(m,l)$, for otherwise the value of inspection $V_q(m) = w_q(m)(1-w_q(m))(X(m,t) - X(m,l))^2 = 0$, violating the receiver's sequential rationality. The sequentially rational action strategy (5) then implies that $X(m,t) > X(m,u) > X(m,l)$ for any $m \in \mathcal{M}_q^+$. If m_q^0 exists, the sender's optimality condition implies that $P(m)u_s(X(m,l)) + (1-P(m))u_s(X(m,u)) = u_s(X(m_q^0, u))$, which means $P(m) = \frac{u_s(X(m,u))-u_s(X(m_q^0,u))}{u_s(X(m,u))-u_s(X(m,l))}$. Since the sender's utility $u_s(.)$ is strictly increasing, there exists such $P(m) \in (0,1]$ if and only if $X(m,l) \leq X(m_q^0, u) < X(m,u)$, which holds for all $m, m' \in \mathcal{M}_q^+$; so, $X(m,l) \leq X(m_q^0, u) < X(m', u)$. If $X(m,l) < X(m_q^0, u)$, it must be $P(m) \in (0,1)$, so sequentially rational inspection requires $V_q(m) = c$. If $X(m,l) = X(m_q^0, u)$, $P(m) = 1$; so, sequentially rational inspection requires $V_q(m) \geq c$.

If m_q^0 does not exist, then the sender's optimality condition implies that for any $m, m' \in \mathcal{M}_q^+$, $P(m)u_s(X(m,l))+(1-P(m))u_s(X(m,u)) = P(m')u_s(X(m',l))+(1-P(m'))u_s(X(m',u)) = u_s(X(m',u))$, which can be achieved with some positive $P(.)$ if and only if $\sup_{m' \in \mathcal{M}_q^+} X(m', l) < \inf_{m'' \in \mathcal{M}_q^+} X(m'', u)$. Sequentially rational inspection requires $V_q(m) = c$ for any m such that $P(m) < 1$, which must be the case when $X(m,l) < \sup_{m' \in \mathcal{M}_q^+} X(m', l)$. $V_q(m) \geq c$ and $P(m) = 1$ is allowed if and only if $X(m,l) = \max_{m' \in \mathcal{M}_q^+} X(m', l)$. ■

The definition and lemma below are useful for proving Proposition 2. For $d \in [2\sqrt{c}, 1]$ and $w^-(.)$ as defined in (15), let

$$h(d) = \frac{w^-(d)}{1 - w^-(d)} \tag{32}$$

be the required liar to truth-teller ratio to maintain incentive for the receiver to inspect a message. It can be verified that $h(.)$ is a strictly decreasing and strictly convex function, with $\lim_{d\to 2\sqrt{c}} h'(d) = -\infty$ and $\lim_{d\to 2\sqrt{c}} h''(d) = +\infty$.

Recall from definition (17) that $X_u^*(x_t, x_l)$ is the induced action when a message is uninspected, where its truth-tellers' expected type is x_t, its liars' expected type is x_l, and its proportion of liars is minimized subject to the constraint $V_q(m) \geq c$. The lemma below shows that $X_u^*(x_t, x_l)$ is increasing in x_t and decreasing in x_l.

Lemma 2 *$\frac{dX_u^*(x_t,x_l)}{dx_t} > 0$ and $\frac{dX_u^*(x_t,x_l)}{dx_l} < 0$.*

Proof of Lemma 2:

$$\begin{aligned}\frac{dX_u^*(x_t,x_l)}{dx_t} =& 1 - w^-(x_t - x_l) - \frac{dw^-(x_t - x_l)}{dx_t}[x_t - x_l] \\ =& \frac{1}{2} + \sqrt{\frac{1}{4} - \frac{c}{(x_t - x_l)^2}} + (\frac{1}{4} - \frac{c}{(x_t - x_l)^2})^{-0.5}\frac{c}{(x_t - x_l)^2} \\ >& 0,\end{aligned}$$

$$\begin{aligned}\frac{dX_u^*(x_t,x_l)}{dx_l} =& w^-(x_t - x_l) - \frac{dw^-(x_t - x_l)}{dx_l}[x_t - x_l] \\ =& \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{c}{(x_t - x_l)^2}} - (\frac{1}{4} - \frac{c}{(x_t - x_l)^2})^{-0.5}\frac{c}{(x_t - x_l)^2} \\ =& (\frac{1}{4} - \frac{c}{(x_t - x_l)^2})^{-0.5}[\frac{1}{2}\sqrt{\frac{1}{4} - \frac{c}{(x_t - x_l)^2}} - (\frac{1}{4} - \frac{c}{(x_t - x_l)^2}) - \frac{c}{(x_t - x_l)^2}] \\ =& (\frac{1}{4} - \frac{c}{(x_t - x_l)^2})^{-0.5}[\frac{1}{2}\sqrt{\frac{1}{4} - \frac{c}{(x_t - x_l)^2}} - \frac{1}{4}] \\ <& 0,\end{aligned}$$

where the last inequality holds because $\sqrt{\frac{1}{4} - \frac{c}{(x_t - x_l)^2}} < \frac{1}{2}$. ■

Proof of Proposition 2:

The reasoning of the proof is as follows: In the most extreme case, an inspected message m can be sent by the truth-tellers with expected type close to 1 and the liars with expected type close to 0. According to the value of inspection derived in (6), the minimum proportion of liars subject to $V_q(m) = c$ for such m would be the smaller solution of $w(1-w)(1-0)^2 = c$, i.e. $w^* = \frac{1}{2} - \sqrt{\frac{1}{4} - c}$. Therefore, the smallest possible expectation of the set of types who send inspected messages is $w^* \times 0 + (1 - w^*) \times 1 = \frac{1}{2} + \sqrt{\frac{1}{4} - c}$.

Since Lemma 1 implies that in any equilibrium, the expectation for the inspected messages (which is $X(m,u)$ for $m \in \mathcal{M}_q^+$) must be higher than the expectation for the uninspected message ((which is $X(m_q^0,u)$), and the expectation of the whole type space $E[\Theta]$ is in between the above two expectation, it must be the case that $E[\Theta] < \frac{1}{2} - \sqrt{\frac{1}{4}-c}$. When this condition is satisfied, we can construct an equilibrium with an inspected message with positive measure sent by a small neighborhood of types below 1 and a small neighborhood of types above 0 and an uninspected message sent by the rest of the types. The formal proof is presented below.

"Only if":

Let $\sigma \equiv (q,P,X)$ be an equilibrium where $p_\sigma > 0$, then the set of inspected messages $\mathcal{M}_q^+$ has a positive measure; therefore, for almost every $m \in \mathcal{M}_q^+$, $X(m,t) = E[\Theta_q^t(m)] < 1$ and $X(m,l) = E[\Theta_q^l(m)] > 0$, so $V_q(m) = w_q(m)(1-w_q(m))(X(m,t)-X(m,l))^2 < \frac{1}{4}$. Since $m \in \mathcal{M}_q^+$ implies that $V_q(m) \geq c$, it must be the case that $c \leq V_q(m) < \frac{1}{4}$.

By the definition of $w^-(.)$ at (15), we have that $w^-(X(m,t)-X(m,l)) = \min\{w \in [0,1] : V_q(m) \geq c\}$. Since for any $m \in \mathcal{M}_q^+$, $V_q(m) \geq c$, so $w_q(m) \geq w^-(X(m,t)-X(m,l))$, and thus $X(m,u) = w_q(m)X(m,l) + (1-w_q(m))X(m,t) \leq w^-(X(m,t)-X(m,l))X(m,l) + (1-w^-(X(m,t)-X(m,l)))X(m,t) \equiv X_u^*(X(m,t),X(m,l))$. Since X is sequentially rational, $X(m,t) = E[\Theta_q^t(m)] \leq 1$ and $X(m,l) = E[\Theta_q^l(m)] \geq 0$, with the inequalities hold strictly if m has a positive measure. Therefore, $X(m,u) \leq X_u^*(X(m,t),X(m,l)) \leq X_u^*(1,0) = \frac{1}{2} + \sqrt{\frac{1}{4}-c}$, where the second inequality holds since by Lemma 2, $X_u^*(X(m,t),X(m,l))$ is strictly increasing in $X(m,t)$ and strictly decreasing in $X(m,l)$, and it holds strictly if m has a positive measure. Recall that $(\Theta, \mathbb{B}(\Theta), \mu)$ is the probability space on which F is defined. Since $\Theta_q^0 \cup \Theta_q^u(\mathcal{M}_q^+) = \Theta$, we obtain $\mu(\Theta_q^0) = 1 - \mu(\Theta_q^u(\mathcal{M}_q^+))$ and

$$(1-\mu(\Theta_q^u(\mathcal{M}_q^+)))E[\Theta_q^0] + \mu(\Theta_q^u(\mathcal{M}_q^+))E[\Theta_q^u(\mathcal{M}_q^+)] = E[\Theta]. \tag{33}$$

Since X is sequentially rational,

$$\begin{aligned}\mu(\Theta_q^u(\mathcal{M}_q^+))E[\Theta_q^u(\mathcal{M}_q^+)] &= \int_{\mathcal{M}_q^+} X(m,u)dH_q(m) \\ &< X_u^*(1,0)\int_{\mathcal{M}_q^+} dH_q(m) \\ &= (\frac{1}{2} + \sqrt{\frac{1}{4}-c})\mu(\Theta_q^u(\mathcal{M}_q^+)),\end{aligned} \tag{34}$$

where the inequality holds strictly since $\mu(\Theta_q^u(\mathcal{M}_q^+)) > 0$. If $\mu(\Theta_q^0) = 0$, then $\mu(\Theta_q^u(\mathcal{M}_q^+)) = 1$ and (33) and (34) imply $E[\Theta] = E[\Theta_q^u(\mathcal{M}_q^+)] < \frac{1}{2} + \sqrt{\frac{1}{4}-c}$. If $\mu(\Theta_q^0) > 0$, then Lemma 1 implies that for

any $m \in \mathcal{M}_q^+$, $E[\Theta_q^0] = X(m_q^0, u) < X(m, u)$, so by (34) $E[\Theta_q^0] < E[\Theta_q^u(\mathcal{M}_q^+)] \leq \frac{1}{2} + \sqrt{\frac{1}{4} - c}$, then (33) implies $E[\Theta] < \frac{1}{2} + \sqrt{\frac{1}{4} - c}$.

"If":

Fix a $\epsilon > 0$, define $\underline{\theta}$ and $\overline{\theta}$ such that they satisfy

$$1 - F(\overline{\theta}) = \epsilon \tag{35}$$

$$F(\underline{\theta}) = h(E[\overline{\theta}, 1] - E[0, \underline{\theta}])\epsilon. \tag{36}$$

Since $c < \frac{1}{4}$, $h(E[\overline{\theta}, 1] - E[0, \underline{\theta}]) \in (0, 1)$ is well defined if $\underline{\theta}$ is close enough to 0 and $\overline{\theta}$ is close enough to 1, therefore $\underline{\theta}$ and $\overline{\theta}$ are well defined for small enough ϵ with $\lim_{\epsilon \downarrow 0} \underline{\theta} = 0$ and $\lim_{\epsilon \downarrow 0} \overline{\theta} = 1$.

Define a message strategy with two on-path messages: the uninspected message $m^0 = (\underline{\theta}, \overline{\theta})$ sent by the set of type $(\underline{\theta}, \overline{\theta})$, and the randomly inspected message $m^+ = [\overline{\theta}, 1]$ sent by the set of type $[0, \underline{\theta}] \cup [\overline{\theta}, 1]$. The corresponding sequentially rational action strategy profile is $X(m^0, u) = E[\underline{\theta}, \overline{\theta}]$, $X(m^+, t) = E[\overline{\theta}, 1]$, $X(m^+, l) = E[0, \underline{\theta}]$ and $X(m^+, u) = w^-(E[\overline{\theta}, 1] - E[0, \underline{\theta}])E[0, \underline{\theta}] + (1 - w^-(E[\overline{\theta}, 1] - E[0, \underline{\theta}]))E[\overline{\theta}, 1]$.

As $\epsilon \to 0$, $\underline{\theta} \to 0$ and $\overline{\theta} \to 1$, therefore $X(m^0, u) \to E[0, 1]$, $X(m^+, l) \to 0$ and $X(m^+, u) \to 1 - w^-(1) = \frac{1}{2} - \sqrt{\frac{1}{4} - c}$, we have $X(m^+, u) > X(m^0, u) > X(m^+, l)$ by assumption 1, thus condition (a) of Lemma 1 is satisfied for small enough ϵ. Denote $d = X(m^+, t) - X(m^+, l)$, then

$$\begin{aligned} w_q(m^+)(1 - w_q(m^+))(X(m^+, t) - X(m^+, l))^2 =& \frac{h(d)}{[1 + h(d)]^2} d^2 \\ =& w^-(d)[1 - w^-(d)]d^2 \\ =& [\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{c}{d^2}}][\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{c}{d^2}}]d^2 \\ =& c. \end{aligned}$$

Thus condition (b) of Lemma 1 is satisfied. Therefore the above message and action strategies profile form an equilibrium with inspection with inspection strategy $P(m^0) = 0$ and $P(m^+) = \frac{u_s(X(m^+, u)) - u_s(X(m^0, u))}{u_s(X(m^+, u)) - u_s(X(m^+, l))}$. Since the receiver is indifferent between inspecting and not inspecting m^+, her equilibrium expected payoff is equivalent to the payoff in which she never inspect m^+ ex-post, therefore according to (13), her equilibrium payoff is $\beta^+(\epsilon)x^+(\epsilon)^2 + (1 - \beta^+(\epsilon))x^0(\epsilon)^2 - E[\theta^2]$, where $\beta^+(\epsilon) \equiv Pr([0, \underline{\theta}] \cup [\overline{\theta}, 1])$, $x^+(\epsilon) \equiv E[[0, \underline{\theta}] \cup [\overline{\theta}, 1]]$ and $x^0(\epsilon) \equiv E[\underline{\theta}, \overline{\theta}]$. The receiver's payoff in the babbling outcome is $E[0, 1]^2 - E[\theta^2]$. Since $\beta^+(\epsilon)x^+(\epsilon) + (1 - \beta^+(\epsilon))x^0(\epsilon) = E[0, 1]$, we have $\beta^+(\epsilon)x^+(\epsilon)^2 + (1 - \beta^+(\epsilon))x^0(\epsilon)^2 - E[0, 1]^2 = \beta^+(\epsilon)(1 - \beta^+(\epsilon))(x^+(\epsilon) - x^0(\epsilon)^2$. Note that $lim_{\epsilon \downarrow 0} x^+(\epsilon) =$

$\frac{1}{2}+\sqrt{\frac{1}{4}-c} > E[0,1] = lim_{\epsilon\downarrow 0} x^0(\epsilon)$; therefore

$$lim_{\epsilon\downarrow 0}\frac{1}{\epsilon}\beta^+(\epsilon)(1-\beta^+(\epsilon))(x^+(\epsilon)-x^0(\epsilon))^2 = (1+h(1))(\frac{1}{2}+\sqrt{\frac{1}{4}-c}-E[0,1])^2 > 0,$$

which implies that there the receiver's payoff from the above equilibrium is higher than her payoff from the babbling outcome for any small enough positive ϵ.

Q.E.D.

Proof of Proposition 3:

We prove this proposition by contradiction. The reasoning is as follows: if there exists a positive measure set of inspected messages such that $V_q(m) > c$, they must be inspected with probability 1, and the liars of these messages induce the same action as the induced action of the uninspected message. Then we can construct a modified equilibrium where a subset of liars who send these inspected messages in the original equilibrium now switch to send the uninspected message. We can do so while maintaining $V_{\hat{q}}(m) = c$ and the same induced action for each type in the modified equilibrium, but a lower ex-ante probability of inspection, thus, a higher ex-ante payoff for the receiver. The formal proof is presented below.

Suppose contrary to the claim, in an optimal equilibrium there exists a positive measure set inspected messages $M_1 \subseteq \mathcal{M}_q^+$ such that $V_q(m) \neq c$ for all $m \in M_1$. Since $P(m) > 0$, sequential rationality (7) then requires that $V_q(m) > c$ and $P(m) = 1$, and (b) of Lemma 1 implies that for all $m \in M_1$, $X(m,l) = \hat{x}$, where $\hat{x} = X(m_q^0, u)$ if m_q^0 exists, $\hat{x} = \max_{m'} X(m', l)$ otherwise. Therefore, we have $E[\Theta_q^l(m)] = E[\Theta_q^l(M_1)] = \hat{x}$ for all $m \in M_1$.

Let $\hat{\Theta} = \Theta_q^l(M_1)$ be the set of liars who send $m \in M_1$ in the original equilibrium. For $m \in M_1$, let $\hat{w}(m) = w^-(X(m,t) - \hat{x})$ be the smallest weight on liars such that value of inspection is no less than c. Since for all $m \in M_1$, $V_q(m) > c$, we have $w_q(m) > \hat{w}(m)$. Now define $\hat{p} = \int_{M_1}(1 - w_q(m))\frac{\hat{w}(m)}{1-\hat{w}(m)}dH_q(m)$, which is the total minimum measure of liars required to match with truth-tellers of $m \in M_1$ such that value of inspection is no less than c. We have $\hat{p} < \mu(\hat{\Theta}) = \int_{M_1} w_q(m)dH_q(m)$.

Assign an arbitrary strict ranking $r : M_1 \to \mathbb{R}$ to the message set M_1. Then for any $m \in M^l$, let

$$z^-(m) = \frac{1}{\mu(\Theta_q^l(M_1))}\int_{m'\in M_1: r(m')<r(m)} \frac{\hat{w}(m')}{1-\hat{w}(m')}(1-w_q(m'))dH_q(m') \tag{37}$$

$$z^+(m) = \frac{1}{\mu(\Theta_q^l(M_1))}\int_{m'\in M_1: r(m')=r(m)} \frac{\hat{w}(m')}{1-\hat{w}(m')}(1-w_q(m'))dH_q(m') \tag{38}$$

be the cumulative required fraction of liars.

For any positive measure set of types $\hat{\Theta}$, define the mean-preserving division $\hat{\Theta}(z) = \hat{\Theta} \cap [\underline{\theta}(z), \overline{\theta}(z)]$ such that $\underline{\theta}(z)$ and $\overline{\theta}(z)$ solve

$$\mu(\hat{\Theta}(z)) = z\mu(\hat{\Theta}) \tag{39}$$

$$E[\hat{\Theta}(z)] = E[\hat{\Theta}]. \tag{40}$$

Define the modified messaging and action strategies $\hat{q}, \hat{X}$ where other things remain unchanged, except the set of messages M_1 and the uninspected message m_q^0. The uninspected message is modified to $m_{\hat{q}}^0 = m_q^0 \cup (\hat{\Theta}/\hat{\Theta}(\frac{\hat{p}}{\mu(\hat{\Theta})}))$, where $\hat{\Theta}/\hat{\Theta}(\frac{\hat{p}}{\mu(\hat{\Theta})})$ is a mean-preserving division of $\hat{\Theta}$ with mean $\hat{x}$ and measure $\mu(\hat{\Theta}) - \hat{p}$. For $m \in M_1$, the set of truth-tellers remain unchanged, while the set of liars is modified to $\Theta_{\hat{q}}^l(m) = \hat{\Theta}(z^+(m))/int(\hat{\Theta}(z^-(m)))$, a mean preserving division of $\hat{\Theta}$ where $int(X)$ is the interior of set X, so that $E[\Theta_{\hat{q}}^l(m)] = \hat{x}$ and the set has measure $\frac{\hat{w}(m)}{1-\hat{w}(m)} \int_{\Theta_q^t(m)} dF(\theta)$.

The sequentially rational actions for the modified uninspected messages $m_{\hat{q}}^0$ is

$$\hat{X}(m_{\hat{q}}^0, u) = E[\hat{\Theta}(\underline{z})] = \hat{x}, \tag{41}$$

and for $m \in M_1$,

$$\begin{aligned} \hat{X}(m,t) &= X(m,t) \\ \hat{X}(m,l) &= X(m,l) = \hat{x} \\ \hat{X}(m,u) &= \hat{w}(m)\hat{x} + (1 - \hat{w}(m))X(m,t), \end{aligned} \tag{42}$$

where $\hat{X}(m,u) > \hat{x}$; thus so $(\hat{q}, \hat{X})$ satisfies (a) of Lemma 1. Furthermore, by the definition of $\Theta_{\hat{q}}^l(m)$ for $m \in M_l$, we have

$$w_{\hat{q}}(m) = \hat{w}(m) = w^-(\hat{X}(m,t) - \hat{X}(m,l)), \tag{43}$$

and

$$V_{\hat{q}}(m) = c. \tag{44}$$

This implies that $(\hat{q}, \hat{X})$ satisfies (b) of Lemma 1. Therefore, there exists $\hat{P}$ such that $\hat{\sigma} = (\hat{q}, \hat{P}, \hat{X})$ is an equilibrium.

Under the modified equilibrium $\hat{\sigma}$, the sequentially rational actions remain unchanged for every type, but the ex-ante probability of inspection is reduced by $\mu(\hat{\Theta}) - \hat{p} > 0$. Therefore, $EU_r(\hat{\sigma}) > EU_r(\sigma)$. This contradicts that σ is an optimal equilibrium.

Q.E.D.

Proof of Proposition 4:

We prove this proposition by contradiction. The reasoning is as follows: if there exists a positive measure set M of inspected messages such that $w_q(m) > 0.5$, we can take a message $m \in M$ with positive measure. From Lemma 1 we have $X(m,l) \leq X(m_q^0, u) < X(m,u) < X(m,t)$, where $X(m,l)$, $X(m,t)$ and $X(m,u)$ are the expected types of liars, truth-tellers and senders of m, and $X(m_q^0)$ is the expected type of the sender of the uninspected message. We then construct a modified equilibrium in which some truth-tellers and liars of m switch to the uninspected message. The sets of truth-tellers and liars of m making this switch satisfy two conditions: (i) the weighted average type that make the switch equals $X(m_q^0, u)$, and (ii) after the switch, the proportion of liars relative to truth-tellers that remain in m becomes $w_{\hat{q}}(m) = 1 - w_q(m)$, so that $V_q(m) = c$ still holds but with minority of liars in m. Since $X(m_q^0, u) < X(m,u)$ in the original equilibrium and a set with expected type equals $X(m_q^0, u)$ switched away from m, the new induced actions $\hat{X}(m,u)$ will be higher than $X(m,u)$ in the modified equilibrium, which means the induced action distribution for the modified equilibrium is a mean-preserving spread of the original one, and by (13) the receiver obtains a higher payoff in the modified equilibrium. The formal proof below handles a more general possibility that allows M to be a set with all zero measure messages that has a positive aggregate measure.

By Proposition 3 $V_q(m) = w_q(m)(1 - w_q(m))(X(m,t) - X(m,l))^2 = c$ for $m \in \mathcal{M}_q^+$, and since $X(m,u) = w_q(m)X(m,l) + (1 - w_q(m)X(m,t))$, we have $(X(m,t) - X(m,u))((X(m,u) - X(m,l))) = c$. Suppose contrary to the claim, in an optimal equilibrium $w_q(m) > 0.5$ for some positive measure set of messages in $\mathcal{M}_q^+$, which implies $X(m,t) - X(m,u) > (X(m,u) - X(m,l))$. Take a positive measure set of message $M^+ \in \mathcal{M}_q^+$ such that $X(m,t) - X(m,u) > (X(m,u) - X(m,l)) + \delta$ for some $\delta > 0$. Then for any $\epsilon > 0$ there exists a positive measure set of message $M_\epsilon^+ \subseteq M^+$ such that for any $m, m' \in M_\epsilon^+$ and $s = t, l, u$, $|X(m,s) - X(m',s)| < \epsilon$ and $X(m,t) - X(m,u) + \delta < X(m,u) - X(m,l)$.

Let $\Theta_\epsilon^l = \Theta_q^l(M_\epsilon^+)$ and $\Theta_\epsilon^t = \Theta_q^t(M_\epsilon^+)$ be the aggregate set of truth-tellers and liars of M_ϵ^+, and $\mu_\epsilon^l = \mu(\Theta_q^l(M_\epsilon^+))$ and $\mu_\epsilon^t = \mu(\Theta_q^t(M_\epsilon^+))$ be the measure of the two sets. Let $E_\epsilon^l = E[\Theta_\epsilon^l]$, $E_\epsilon^t = E[\Theta_\epsilon^t]$ and $E_\epsilon^u = E[\Theta_\epsilon^l \cup \Theta_\epsilon^t]$ be the corresponding expected values of the sets. Note that $|E_\epsilon^s - X(m,s)| < \epsilon$ for any $m \in M_\epsilon^+$ and $s = t, l, u$; thus, we have

$$E_\epsilon^t - E_\epsilon^u > E_\epsilon^u - E_\epsilon^l + \delta - 2\epsilon \tag{45}$$

$$|(E_\epsilon^t - E_\epsilon^u)(E_\epsilon^u - E_\epsilon^l) - c| < 4\epsilon. \tag{46}$$

Let $\hat{E}$ be the larger root of $(E_\epsilon^t - \hat{E})(\hat{E} - E_\epsilon^l) - c = 0$. (45) and (46) imply that for small enough ϵ, $E_\epsilon^t - \hat{E} < \hat{E} - E_\epsilon^l$ and $\hat{E} > E_\epsilon^u + \delta$. Fix any $m \in \mathcal{M}_q^+$, and let $\underline{u} = P(m)u_s(X(m,l)) + (1 - P(m))u_s(X(m,u))$ be the expected payoff of the liars and $\hat{x} = u^{-1}(\underline{u})$ be its certainty equivalence. Note that Proposition 1 implies any liar can mimic the payoff of any other liar, which means all liars receive the same payoff in the equilibrium, and $\hat{x} = X(m_q^0, u)$ if an uninspected message m_q^0 exists. Lemma 1 implies $X(m,l) \leq \hat{x} < X(m,u)$ for any $m \in M_\epsilon^+$; thus, we have

$$E_\epsilon^l \leq \hat{x} < E_\epsilon^u < \hat{E} - \delta < E_\epsilon^t - \delta. \tag{47}$$

Let z_l, z_t solve

$$z_l \mu_\epsilon^l E_\epsilon^l + z_t \mu_\epsilon^t E_\epsilon^t = (z_l \mu_\epsilon^l + z_t \mu_\epsilon^t)\hat{x} \tag{48}$$

$$(1 - z_l)\mu_\epsilon^l E_\epsilon^l + (1 - z_t)\mu_\epsilon^t E_\epsilon^t = [(1 - z_l)\mu_\epsilon^l + (1 - z_t)\mu_\epsilon^t]\hat{E}. \tag{49}$$

Since $\mu_\epsilon^l E_\epsilon^l + \mu_\epsilon^t E_\epsilon^t = (\mu_\epsilon^l + \mu_\epsilon^t)E_\epsilon^u$, so (47) means $z_l \in (0,1)$ and $z_t \in [0,1)$

Recall the mean-preserving division we defined in (39) and (40). We divide the liar set Θ_ϵ^l into $\Theta_\epsilon^l(z_l)$ and $\Theta_\epsilon^l / \Theta_\epsilon^l(z_l)$, and truthful set Θ_ϵ^t into $\Theta_\epsilon^t(z_t)$ and $\Theta_\epsilon^t / \Theta_\epsilon^t(z_t)$. The mean-preserving divisions imply $E[\Theta_\epsilon^l(z_l)] = E[\Theta_\epsilon^l / \Theta_\epsilon^l(z_l)] = E_\epsilon^l$ and $E[\Theta_\epsilon^t(z_t)] = E[\Theta_\epsilon^t / \Theta_\epsilon^t(z_t)] = E_\epsilon^t$. From (48) and (49) we have $E[\Theta_\epsilon^l(z_l) \cup \Theta_\epsilon^t(z_t)] = \hat{x}$ and $E[\Theta_\epsilon^l / \Theta_\epsilon^l(z_l) \cup \Theta_\epsilon^t / \Theta_\epsilon^t(z_t)] = \hat{E}$.

Now define a modified equilibrium $\hat{\sigma} = (\hat{q}, \hat{P}, \hat{X})$ where other things remain unchanged, except the set of messages M_ϵ^+ is off-path and a message $\hat{m} = \Theta_\epsilon^t / \Theta_\epsilon^t(z_t)$ is added with $\hat{q}(\theta) = \hat{m}$ for $\theta \in \Theta_\epsilon^l / \Theta_\epsilon^l(z_l) \cup \Theta_\epsilon^t / \Theta_\epsilon^t(z_t)$. The uninspected message m_q^0 (if exists) is modified to $m_{\hat{q}}^0 = m_q^0 \cup \Theta_\epsilon^l(z_l) \cup \Theta_\epsilon^t(z_t)$ with $\hat{q}(\theta) = m_{\hat{q}}^0$ for $\theta \in \Theta_\epsilon^l(z_l) \cup \Theta_\epsilon^t(z_t)$.

The sequentially rational actions for the modified messages $\hat{m}$ and $m_{\hat{q}}^0$ are $\hat{X}(\hat{m}, t) = E_\epsilon^t$, $\hat{X}(\hat{m}, l) = E_\epsilon^l$, $\hat{X}(\hat{m}, u) = \hat{E}$, and $\hat{X}(m_{\hat{q}}^0, u) = \hat{x}$. By (47) we still have $\hat{X}(m,l) \leq \hat{X}(m_{\hat{q}}^0, u) < \hat{X}(m,u)$ for all $m \in \mathcal{M}_{\hat{q}}^+$; thus, (a) in Lemma 1 is satisfied. For the newly added inspected message $\hat{m}$, $(\hat{X}(m_{\hat{q}}^0, t) - \hat{X}(m_{\hat{q}}^0, u))(\hat{X}(m_{\hat{q}}^0, u) - \hat{X}(m_{\hat{q}}^0, l)) = (E_\epsilon^t - \hat{E})(\hat{E} - E_\epsilon^l) = c$; thus, (b) in Lemma 1 is satisfied. Therefore there exists $\hat{P}$ such that $\hat{\sigma}$ is an equilibrium.

To compare the receiver's exante payoffs, for any equilibrium σ, let

$$G_\sigma^u(x) = \int_{\mathcal{M}_q} \mathbf{1}(X(m,u) \leq x) dH_q(m)$$

be the distribution of induced **uninspected** actions under σ. Since Proposition 3 implies $V_q(m) = c$ for any inspected messages in an optimal equilibrium, the receiver is indifferent between inspecting

and not inspecting any $m \in \mathcal{M}_q^+$. Therefore, it is equivalent to express the receiver's expected payoff as if m is not inspected ex-post. The above argument combined with the receiver's payoff function (13) implies that in any optimal equilibrium, $EU_r(\sigma) = \int_0^1 x^2 dG_\sigma^u(x) - E(\theta^2)$. Let G_σ^u and $G_{\hat{\sigma}}^u$ be the distribution of induced uninspected actions of the two equilibria. By sequential rationality the two distributions have the same mean $\int_0^1 x dG_{\hat{\sigma}}^u(x) = \int_0^1 x dG_\sigma^u(x) = \int_\Theta \theta dF(\theta)$ and they differ only by actions induced by the set $\Theta_\epsilon^l \cup \Theta_\epsilon^t$. In the original equilibrium σ, a type in $\Theta_\epsilon^l \cup \Theta_\epsilon^t$ sends some $m \in M_\epsilon^+$ with induced action $X(m,u)$ where $|X(m,u) - E_\epsilon^u| < \epsilon$; in the modified mechanism $\hat{\sigma}$, a type in $\Theta_\epsilon^l \cup \Theta_\epsilon^t$ sends either $\hat{m}$ or $m_{\hat{q}}^0$ with induced action either $\hat{X}(m_{\hat{q}}^0, u) = \hat{x}$ or $\hat{X}(\hat{m}, u) = \hat{E}$. (47) implies that for small enough ϵ, $X(m_{\hat{q}}^0, u) < X(m,u) < \hat{X}(\hat{m}, u)$ for any $m \in M_\epsilon^+$. Therefore, $G_{\hat{\sigma}}^u$ is a mean-preserving spread of G_σ^u, which means $\int_0^1 x^2 dG_{\hat{\sigma}}^u(x) > \int_0^1 x^2 dG_\sigma^u(x)$, then (13) implies $EU_r(\hat{\sigma}) > EU_r(\sigma)$. This contradicts that σ is an optimal equilibrium.

Q.E.D.

Proof of Proposition 5:

Recall that $\Theta_q^0 = \{\theta : P(q(\theta)) = 0)\}$ is the set of uninspected types in the equilibrium $\sigma = (q, P, X)$. Since X is sequentially rational, $X(q(\theta), u) = E[\Theta_q^0]$ for any $\theta \in \Theta_q^0$. Therefore, the receiver's equilibrium payoff for the set of uninspected types is

$$-\int_{\Theta_q^0} (\theta - E[\Theta_q^0])^2 d\theta. \tag{50}$$

Let $\alpha = \int_{\Theta_q^0} d\theta$ be the probability of uninspected types in equilibrium. Clearly, (50) is maximized when Θ_q^0 is an interval with length α. Given the uniform distribution, the payoff is $-\frac{\alpha^3}{12}$.

Let $\Theta_q^+ = \{\theta : P(q(\theta) > 0)\}$ be the set of inspected types in the equilibrium. For any m such that $P(m) > 0$, the receiver either strictly prefers inspecting to not inspecting or she is indifferent. Therefore, for any m such that $P(m) > 0$, It is equivalent to express the receiver's expected payoff as if m is inspected ex-post. As a result, the receiver's equilibrium payoff for the set of inspected types can be written as

$$-\int_{\Theta_q^+} [(\theta - X(q(\theta), s(\theta)))^2 + c] d\theta, \tag{51}$$

where $s(\theta) = t$ if $\theta \in q(\theta)$ (truth-teller) and $s(\theta) = l$ if $\theta \notin q(\theta)$ (liar). (51) is maximized when

$X(q(\theta), s(\theta)) = \theta$ for all $\theta \in \Theta_q^+$, i.e. perfect information upon inspection. In that case, the payoff is $-c(1-\alpha)$.

The function $g(\alpha) = -\frac{\alpha^3}{12} - c(1-\alpha)$ is strictly concave with a unique maximum at $\alpha = 2\sqrt{c} \leq 1$ for any $c \leq \frac{1}{4}$. Therefore, for any equilibrium (q, P, X), $EU_r(q, P, X) \leq g(2\sqrt{c})$.

Q.E.D.

Proof of Proposition 6:

To show that σ_d is an equilibrium, for $m \in \mathcal{M}_q^+$,

$$\begin{aligned}\frac{w_q(m)}{1-w_q(m)} &= \lim_{\epsilon \to 0} \frac{\mu([\phi_d(m+\epsilon), \phi_d(m-\epsilon)])}{\mu([m-\epsilon, m+\epsilon])} \\ &= -\dot{\phi}_d(m) = \frac{w^-(m-\phi_d(m))}{1-w^-(m-\phi_d(m))},\end{aligned} \tag{52}$$

where the first equality holds because of the continuously decreasing message strategy, and the third equality holds by (19). Therefore, $w_q(m) = w^-(m-\phi_d(m))$ and $V_q(m) = c$, thus condition (b) of Lemma 1 is satisfied.

$X(m_q^0, u)$, $X(m, t)$ and $X(m, l)$ for $m \in \mathcal{M}_q^+$ are clearly sequentially rational given the message strategy. For $m \in \mathcal{M}_q^+$,

$$\begin{aligned}X(m,u) = X_u^*(m, \phi_d(m)) &= w^-(m-\phi_d(m))\phi_d(m) + (1-w^-(m-\phi_d(m)))m \\ &= w_q(m)X(m,l) + (1-w_q(m))X(m,t)\end{aligned}$$

is also sequentially rational. For any $m \in \mathcal{M}_q^+ = (\overline{\theta}_d, 1]$,

$$\begin{aligned}X(m,u) = X_u^*(m, \phi_d(m)) > E[\phi_d(m), m] &> E[\phi(\overline{\theta}_d), \overline{\theta}_d] \\ &= X(m_q^0, u) > \underline{\theta}_d > \phi_d(m) = X(m,l),\end{aligned}$$

where the first inequality holds because $X_u^*(m, \phi_d(m)) = w^-(m-\phi_d(m))\phi_d(m) + (1-w^-(m-\phi_d(m)))m > \frac{\phi_d(m)+m}{2}$ as $w^-(m-\phi_d(m)) < 0.5$; the second inequality holds because by differential equation (19), $\frac{m-\overline{\theta}_d}{\phi(\overline{\theta}_d)-\phi(m)} = \int_{\overline{\theta}_d}^m \frac{1-w^-(x-\phi_d(x))}{w^-(x-\phi_d(x))}dx > 1$, so $\frac{m+\phi_d(m)}{2} > \frac{\overline{\theta}_d+\underline{\theta}_d}{2}$. Therefore, condition (a) of Lemma 1 is satisfied, and thus σ_d is an equilibrium with the inspection strategy specified by (24).

Since $X(m,u) = X_u^*(m, \phi_d(m))$, and $\frac{dX_u^*(m,\phi_d(m))}{dm} = \frac{\partial X_u^*(m,\phi_d(m))}{\partial m} + \frac{\partial X_u^*(m,\phi_d(m))}{\partial \phi_d(m)}\dot{\phi}_d(m) > 0$ because $\dot{\phi}_d(m) < 0$ and by Lemma 2 $\frac{\partial X_u^*(m,\phi_d(m))}{\partial m} > 0$ and $\frac{\partial X_u^*(m,\phi_d(m))}{\partial \phi_d(m)} < 0$. Therefore, $X(m,u)$ is increasing in m.

To show that σ_d is an optimal equilibrium, the set of uninspected set in σ_d is $\Theta_q^0 = (\underline{\theta}_d, \overline{\theta}_d)$, which is an interval of length $2\sqrt{c}$. For all $m \in [\overline{\theta}_d, 1]$, $X(m,t) = m$ and $X(m,l) = \phi(m)$; therefore, all inspected types are revealed upon inspection. Therefore, $EU_r(\sigma_d) = -\frac{\alpha^3}{12} - c(1-\alpha)$, where $\alpha = 2\sqrt{c}$, which is the upper bound of the receiver's equilibrium payoff shown in Proposition 5.

Q.E.D.

Proof of Proposition 7:

(i) We claim that for any $m, m' \in \mathcal{M}_q$ such that $P(m) < 1$ and $P(m') < 1$, $X(m,u) = X(m',u)$ and $P(m) = P(m')$. Suppose contrary to the claim, $X(m',u) > X(m,u)$. Then since $X(m,u) = E[\Theta_q^u(m)]$, there exists $\overline{\theta} \geq X(m,u)$ who sends m in the equilibrium. The optimality condition of type $\overline{\theta}$ implies $[1-P(m)][u_s(X(m,u)) - u_s(\overline{\theta})] \geq [1-P(m')][u_s(X(m',u)) - u_s(\overline{\theta})]$, thus $P(m) > P(m')$. Similarly, there exists $\underline{\theta} \leq X(m,u)$ who sends m in the equilibrium. The optimality condition of type $\underline{\theta}$ implies $[1-P(m)][u_s(X(m,u)) - u_s(\underline{\theta})] \geq [1-P(m')][u_s(X(m',u)) - u_s(\underline{\theta})]$, thus $P(m) < P(m')$. The above two results contradicts each other. Therefore, it must be the case that $X(m,u) = X(m',u)$. Given this result, optimality of the sender of m implies $1-P(m) \geq 1-P(m')$ while optimality of the sender of m' implies $1-P(m') \geq 1-P(m)$. Therefore, $P(m) = P(m')$.

(ii) We claim that if there exists $\tilde{m} \in \mathcal{M}_q$ such that $P(\tilde{m}) = 1$, then $P(m) = 1$ for almost every $m \in \mathcal{M}_q$. Suppose contrary to the claim, there exists a positive measure subset $\hat{\Theta} \subseteq \Theta$ such that $P(q(\theta)) < 1$ for any $\theta \in \hat{\Theta}$, then the claim in (i) implies that $X(q(\theta),u) = E[\hat{\Theta}]$ for any $\theta \in \hat{\Theta}$. Since $\mu(\hat{\Theta}) > 0$, there exists $\theta' \in \hat{\Theta}$ such that $\theta' > E[\hat{\Theta}] = X(q(\theta'),u)$, but then type θ' receives $P(q(\theta'))u_s(\theta') + (1-P(q(\theta')))u_s(X(q(\theta'),u)) < u_s(\theta')$ in the equilibrium, which is not optimal as he would receive $u_s(\theta')$ if he deviates to $\tilde{m}$, and this proves the claim.

The above two claims imply that $P(m) = \hat{P}$ for almost every $m \in \mathcal{M}_q$, where $\hat{P}$ is a constant, and if $\hat{P} < 1$, $X(m,u) = E[\Theta]$ for almost every $m \in \mathcal{M}_q$. By the law of total variance, $Var(\Theta) \geq \int_{\mathcal{M}_q} Var(\Theta_q^u(m))dH_q(m)$. Therefore, if $c > Var(\Theta)$, then there exists some equilibrium messages m such that $c > Var(\Theta) \geq Var(\Theta_q^u(m)) = V_q(m)$, thus, sequential rationality of P implies $\hat{P} = P(m) = 0$. To show that if $c < Var(\Theta)$ then $\hat{P} = 1$, suppose on the contrary $c < Var(\Theta)$ and $\hat{P} < 1$. Since $X(m,u) = E[\Theta]$ for almost every equilibrium message m, $\int_{\mathcal{M}_q} Var(\Theta_q^u(m))dH_q(m) = \int_\Theta (\theta - E[\Theta])^2 dF(\theta) = Var(\Theta)$. Therefore, there exists some equilibrium messages m such that $c < Var(\Theta) \leq Var(\Theta_q^u(m)) = V_q(m)$, contradicting $\hat{P} < 1$. We conclude that if $c > Var(\Theta)$,$P(m) = 0$ for almost every $m \in \mathcal{M}_q$, and the equilibrium is uninformative; if $c < Var(\Theta)$, then $c < V_q(m)$ and

$P(m) = 1$ for almost every $m \in \mathcal{M}_q$, the equilibrium is state-verifying.

Q.E.D.

Proof of Proposition 8:

Suppose $c \geq Var(\Theta)$, then $EU_r^v = Var(\Theta)$. Consider a strategy profile σ under lie-detecting technology, where $m_q(\theta) = m_q^0 = \Theta$ for any $\theta \in \Theta$ and $P(m_q^0) = 0$, $X(m_q^0, u) = E[\Theta]$. It is clear that such strategy profile is an uninformative equilibrium, and $EU_r(\sigma) = Var(\Theta) = EU_r^v$. Therefore, $EU_r(\sigma^*) \geq EU_r(\sigma) = EU_r^v$.

Suppose $c < Var(\Theta)$, then $c < Var(\Theta) < (1 - E\Theta)(E(\Theta) - 0) \leq \frac{1}{4}$, where the second inequality holds by Bhatia–Davis inequality. Since $c < (1 - E\Theta)(E(\Theta) - 0)$, we have $\frac{1}{2} - \sqrt{\frac{1}{4} - c} < E(\Theta) < \frac{1}{2} + \sqrt{\frac{1}{4} - c}$, thus Assumption 1 is satisfied, and by proposition 6, σ_d is an equilibrium with $EU_r(\sigma_d) = -\frac{(2\sqrt{c})^3}{12} - c(1 - 2\sqrt{c}) > -c$. Therefore, $EU_r(\sigma^*) = EU_r(\sigma_d) > -c = EU_r^v$.

Q.E.D.