Communication with Forgetful Liars*

Philippe Jehiel†

28th August 2020

Abstract

I consider multi-round cheap talk communication environments in which, after a lie, the informed party has no memory of the content of the lie. I characterize the equilibria with forgetful liars in such settings assuming that a liar's expectation about his past lie coincides with the equilibrium distribution of lies aggregated over all possible realizations of the states. The approach is used to shed light on when the full truth is almost surely elicited, and when multiple lies can arise in equilibrium. Elaborations are proposed to shed light on why non-trivial communication protocols are used in criminal investigations.

Keywords: forgetful liars, lie detection, analogy-based expectations, cheap talk

*I wish to thank Thomas Mariotti (the editor) and three anonymous referees for constructive comments. I also that Johannes Hörner, Navin Kartik, Frédéric Koessler, Joel Sobel, Rani Spiegler as well as seminar participants at PSE, Warwick theory workshop, Barcelona workshop, Glasgow university, Lancaster game theory workshop, D-Tea 2018, University of Bonn, ESSET 2018, and Stockholm University for useful comments. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 742816).

†PSE, 48 boulevard Jourdan, 75014 Paris, France and University College London ; jehiel@enpc.fr

1 Introduction

In criminal investigations, it is of primary importance to detect when a suspect is lying. Quite commonly, suspects are requested to tell an event several times, possibly in different frames, and inconsistencies across the reports are typically used to detect lies, and obtain admission of guilt. As formulated in Vrij et al. (2011), the benefit of repeating the request is that *a liar's memory of a fabricated answer may be more unstable than a truth-teller's memory of the actual event.* As a result, it may be harder for a lying suspect than for a truth-teller to remain consistent throughout, which can then be exploited by investigators.

Such a view about the potential instability in liars' memory has been investigated experimentally by a number of scholars typically outside economics (see the discussion and literature review in Vrij et al. (2011)). The objective of this paper is to develop a game theoretic framework and corresponding solution concept that formalize it. Specifically, I am interested in understanding how the asymmetry in memory between liars and truth-tellers can affect the strategy of communication of informed parties. To this end, I consider standard communication settings in which there is a conflict of interest between an informed party (denoted I) who knows an event s and an uninformed party (denoted U) who does not know s but would like to learn about it. Communication about s takes place in more than one round so that there is room for a liar to forget some of what he previously said.

Key questions of interest are: Does the informed party engage into lying, and if so in what kind of events s and with what kind of lies? Do inconsistencies trigger harmful consequences? Are there circumstances in which the full truth about the event is almost surely elicited?

Addressing such questions is of clear interest to the understanding of any strategic communication setting to the extent that the memory asymmetry between liars and truth-tellers seems widespread.[1] An important game theoretic insight obtained for such interactions in the absence of memory imperfections has been that full information transmission should not be expected, as soon as there are conflicts of interest (Crawford and

[1] To substantiate this, it may be relevant to mention a popular quote attributed to Mark Twain " *When you tell the truth you do not have to remember anything,"* which subtly suggests a memory asymmetry whether you tell the truth or you lie.

Sobel (1982)). But, how is this insight affected when liars are forgetful?

A key modeling choice concerns the expectations of liars with respect to the content of their past lies. I will have in mind environments in which a given individual in the role of party I would not engage himself very often in the communication game. Thus, he would not know how he (routinely) communicates as a function of s. However, he would know, through learning from others' experiences, the empirical distribution of lies (as aggregated over different realizations of s). I will be assuming that when party I lies, he later believes he used a communication strategy that matches this aggregate empirical distribution.[2]

To state the main insights, let me complete the description of the communication setting. The events referred to as sates s can take discrete values in $S \subseteq [0,1]$, and each realization of s can occur with a probability known to party U. In the criminal investigation application, the various s correspond to different levels of guilt where $s = 1$ can be interpreted as complete innocence and $s = 0$ as full guilt. After hearing party I, party U chooses the action that matches her expectation of the mean value of s, an action that affects party I 's well-being.

Communication does not take place at just one time. Specifically, two messages m_1 and m_2 are being sent by party I at two different times $t = 1, 2$. If party I in state s tells the truth by communicating $m_1 = s$ at time $t = 1$, he remembers it at time $t = 2$, but if he lies by communicating $m_1 \neq s$, he does not remember at time $t = 2$ what message was sent at time $t = 1$.[3] He is always assumed to know the state s though. That is, the imperfect memory is only about the message sent at time $t = 1$, not about the state. Party U is assumed to make the optimal choice of action given the messages (m_1, m_2) she receives.

As highlighted above, I assume that when party I lies at $t = 1$, he believes at $t = 2$ that he sent a message at $t = 1$ that matches the aggregate distribution of lies as

[2]Such an assumption implicitly requires that previous messages not corresponding to the truth are disclosed and tagged as being lies before the details of the state are disclosed, thereby ensuring that there is no access to the joint distribution of previous messages and states. See below for elaborations on this where I also discuss alternative specifications.

[3]My approach thus assumes that messages have an accepted meaning so that lying can be identified with sending a message that differs from the truth (see Sobel (2018) for a recent contribution that provides a definition of lying in communication games that agrees with this view).

occurring in equilibrium across the various states. All other expectations of party I are assumed to be correct, and strategies are required to be best-responses to expectations, as usual. The corresponding equilibria are referred to as equilibria with forgetful liars. I characterize such equilibria in the communication setting just described adding the (small) perturbations that, party I incurs a tiny extra cost when lying, and, with a tiny probability, party I always communicates the truth.[4] The main findings are as follows.

I first consider pure persuasion situations in which party I 's objective is the same for all states and consists in inducing a belief about s as high as possible in party U's mind. For such specifications, the equilibria employing pure strategies have the following form. Either party I always tells the truth or there is exactly one lie made in equilibrium. In the latter case, the unique lie s^h belongs to the state space S, and party I chooses to lie when the state s is below a threshold s^l defined so that $E(s \in S,\ s \leq s^l \text{ or } s = s^h)$ is in between s^l and the state in $S \setminus \{s^h\}$ just above s^l. Moreover, when considering the fine grid case in which two consecutive states are close to each other and all possible states can arise with a probability of similar magnitude, I show that all pure strategy equilibria with forgetful liars lead approximately to the first-best in which party U perfectly infers the state whatever s, and chooses the action $a = s$ accordingly.

The reason why some one-lie communication strategies can be sustained as equilibria is that then the aggregate distribution of lies is concentrated on just one realization so that a liar by making this common lie can ensure he will not be caught being inconsistent. Arguments similar to the unravelling argument are next used to complete the characterization of pure strategy equilibria.

I also briefly discuss a class of mixed strategy equilibria,[5] and observe for those that multiple lies can occur, inconsistent messages can happen leading to less good outcomes for party I, and, as for the pure strategy equilibria, the first-best is asymptotically approached in the fine grid case.

[4] These perturbations ensure that if party I is indifferent between lying and truth-telling, he chooses truth-telling, and if off-the-path messages $m_1 = m_2 = s \in S$ were received, party U would believe the state is s. I will also assume that when off-the-path message profiles other than (s, s) are received, the action chosen by party U is 0 (or small enough), which is required to support pure strategy equilibria (see discussion below).

[5] These can be shown to be the only ones under additional perturbations (see the working papr version Jehiel (2019)).

Thus, in pure persuasion situations, when liars are forgetful, simple multi-round communication protocols ensure that party U obtains much more information transmission from party I as compared with one-shot communication protocols in which party I would not disclose any information. Moreover, when there is some significant lying activity (i.e. moving away from the fine grid case), there is only one lie occurring in pure strategy equilibria, this unique lie is made only for low levels of s, and party I is never caught making inconsistent lies.

I next explore the effect of letting the objective of the informed party I depend also on the state s. The main observation in this case is that multiple lies can sometimes arise in equilibria employing pure strategies. The reason is as follows. After a lie at $t = 1$, even though, irrespective of s, party I at $t = 2$ holds the same belief about the message sent at $t = 1$, he may now opt for different m_2 depending on the state s because party I rightly understands how party U's action varies with the messages and party I's payoff depends on s, unlike in the pure persuasion case. This observation can be used to construct equilibria in which depending on the state, liars find it strictly beneficial to sort into different lies without ever being inconsistent.

In the final part of the paper, I briefly consider an extension (with the criminal investigation application in mind) in which the state takes a more complex form with two attributes s_A and s_B whose sum $s = s_A + s_B$ determines the level of guilt, and the imperfect memory of a liar concerns the details describing the lie (the exact profile of reported attributes) but not the targeted level of guilt (as represented by the sum of the reported attributes). When the communication protocol takes a sufficiently non-trivial form (with randomization on the order in which the details are requested at $t = 1$ and randomization on which attribute is requested at $t = 2$), the equilibrium outcomes of the communication game with forgetful liars (to be extended appropriately) are very similar to the ones arising in the basic model (with only one lie being made in the pure strategy equilibria in the pure persuasion scenario and almost perfect information elicitation in the fine grid case). Interestingly, more equilibrium outcomes (including ones which are bounded away from the first-best in the fine grid case) can be supported if the communication protocol is too simple (for example as resulting from protocols in which at $t = 2$, party I is always asked to report the realization of the same pre-specified attribute). Such additional insights

while obtained in a stylized model can be viewed as shedding light on why non-trivial communication protocols are generally used in criminal investigations (see also Vrij et al. (2008) for experimental finding showing the benefit of increasing the cognitive load for communication transmission purposes).

Related Literature

The paper can be related to different strands of literature. First, there is a large literature on cheap talk initiated by Crawford and Sobel (1982) (see also Green and Stokey (2007)), which has emphasized that in the presence of conflicts of interest, some information would be withheld by the informed party. The analysis of strategic communication is deeply affected by the forgetfulness of liars, as exemplified by the almost perfect information transmission obtained in a two-round communication protocol when the grid of states is fine. In relation to the cheap talk literature, it should be mentioned that while most of this literature has considered one-round communication protocols, it has also observed that with multiple rounds, more equilibrium outcomes can be supported. The logic of such results is however unrelated to the memory imperfections considered in this paper, and for example the almost perfect elicitation of the state would never arise even with multiple rounds when there is no memory imperfection.[6]

Second, the equilibria with forgetful liars turn out to be similar to the Perfect Bayesian Nash equilibria that would arise in certification games in which all types but those corresponding to the lies could certify all what they know (see Grossman and Hart (1980), Grossman (1981), Milgrom (1981), Dye (1985) or Okuno-Fuijiwara and Postlewaite (1990) for some key references in the certification literature).[7] In particular, when there is only one lie s^h as in the pure strategy equilibria of the pure persuasion games, the equilibrium

[6] With perfect memory, multi-round communication protocols allow to implement a larger spectrum of the communication equilibria that could be obtained through the use of a mediator as compared with the smaller set of Nash equilibria that can be implemented with one round of direct communication between the two parties (see Forges (1990), Aumann and Hart (2003) or Krishna and Morgan (2004) for discussion of this).

[7] Interestingly, Mark Twain's quote as reported in footnote 1 has sometimes been used to motivate that explicit lies (as opposed to lies by omission) may be costly or simply impossible as in certification games (see, for example, Hart, Kremer and Perry (2017)). By contrast, my approach can be viewed as offering an explicit formalization of memory asymmetry between liars and truth-tellers as suggested in that quote. It may be mentioned here that the same Twain quote appears also in a recent paper by Hörner et al. (2017) on dynamic communication with Markovian transitions between states, but the link to the present study in which there is no evolution of states is even less immediate.

outcome is similar to that in Dye (1985)'s model identifying type s^h in my model with the type that cannot be certified (the uninformed type) in his. Of course, a key difference is that, in this analogy, the set of types that cannot be certified is not exogenously given in the present context, as it is determined by the set of lies made in equilibrium, which is endogenously determined.

Third, the proposed modeling of the expectation of a forgetful liar is in the spirit of the analogy-based expectation equilibrium ((Jehiel (2005) and Jehiel and Koessler (2008)) to the extent that the considered distribution of messages is the overall distribution of lies aggregated over all states, and not the corresponding distribution conditioned by the state. I briefly discuss below the case in which a forgetful liar would use the conditional distribution instead (this alternative modeling would be in the spirit of either of the multi-selves approaches considered by Piccione and Rubinstein (1997) and fits applications in which party I would know how his lying strategy varies with s for example because he would have played himself the game many times). I note that with such a modeling, many more equilibrium outcomes can be supported. In particular, in the pure persuasion case, I construct such pure strategy equilibria the outcome of which is far away from the first-best, even in the fine grid case.

Fourth, putting the present paper in the perspective of other behavioral models of strategic communication, I note that a number of these consider the modeling of deception, which is concerned with how the informed party can manipulate the belief of the uninformed party (a behavioral dimension not present here). These include Crawford (2003) who adds in a game in which players communicate about their intended action the possibility that players interpret the declared intention naively as in the level-k approach, Kartik et al. (2007) who consider one-shot cheap talk games with unbounded state space again when some share of receivers interpret naively what they are told (with the observation that the cheap talk game is then transformed into a signalling game admitting a separating equilibrium when the state space is unboundedness) or Ettinger and Jehiel (2010) who consider a different application of the analogy-based expectation equilibrium to communication games, this time focused on the coarse understanding of the uninformed party rather than the memory limitations of the informed party. Concerning behavioral twists on the informed party' side, one may mention the work of Kartik (2009)

who adds explicit (and not vanishingly small, as considered here) lying costs to the standard cheap talk game, and observe in a setup with bounded state space that every type has an incentive to inflate his type with some pooling at the highest messages (which sharply contrasts with the shape of the equilibria with forgetful liars in pure persuasion situations as described above in which pooling occurs for low types) or the work of Deneckere and Severinov (2017) with similar lying costs this time considered in more flexible mechanism design settings.[8]

Finally, it may be worth mentioning the work of Dziuda and Salas (2018) who consider one-round communication settings similar to those in Crawford and Sobel in the pure persuasion game scenario (see also Balbuzanov (2017) for the case of state-dependent preferences) in which a lie made by the Sender may sometimes be detected by the Receiver. Thinking of the observation of inconsistencies by the uninformed party as a lie detection technology, it would seem the present paper proposes an endogenous channel through which lies are detected. Yet, this is not the driving force behind the analysis here as in many equilibria with forgetful liars (in particular those employing pure strategies), there is no inconsistency in equilibrium and thus no lie detection as in Dziuda and Salas (it is rather the fear of being inconsistent if lying that drives the equilibrium choice of strategy of the informed party).[9]

The rest of the paper is organized as follows. Section 2 describes the model and solution concept. Section 3 analyzes pure persuasion situations. Section 4 analyzes a simple class of state-dependent preferences. Section 5 offers a discussion. Section 6 concludes.

[8] Deneckere and Severinov (2017) assume that each time the informed party misreports his type, he incurs an extra cost. They observe that using multiround mechanisms (in which if consistently lying the informed party would have to incur prohibitive cost) may help extract the private information at no cost. While the benefit of multiround communication is common to my approach and theirs, the main contribution of the present study concerns the endogenous derivation of lying costs as arising from memory limitations in given communication games. This is clearly complementary to the mechanism design perspective of their approach in which lying costs are exogenously given.

[9] Clearly, the informational settings are very different in the two papers: there is no memory issue on the informed party side in Dziuda and Salas and there is no technology for lie detection in my setting. Yet, a common feature of the analysis is that Senders in favorable states prefer telling the truth. But, note that the shape of the lying strategy of those senders in unfavorable states is different as these randomize over a full range of messages above a threshold in Dziuda and Salas, which is not so in my setting.

2 The Model

Events s -referred to as states- can take n possible values $s_1 < s_2 < ...s_n$ with $s_1 = 0$ and $s_n = 1$ where $S = \{s_k\}_{k=1}^n$ denotes the state space. The ex ante probability that state s_k arises is $p(s_k)$, which is commonly known. There are two parties, an informed party I and an uninformed party U. The informed party knows the realization of the state $s \in S$, the uninformed party does not.

Party I first communicates about s according to a protocol to be described shortly. At the end of the communication phase, party U has to choose an action $a \in [0, 1]$. The objective of party U takes the quadratic form $-(a - s)^2$ so that she chooses the action a that corresponds to the expected value of s given what she believes about its distribution.

Party I cares about the action a chosen by U and possibly (but not necessarily) about the state s. Ignoring for now the messages sent during the communication phase, party I's payoff can be written as $u(a, s)$.

I will start the analysis with pure persuasion situations in which party I would like the action a to be as large as possible independently of s. I will next discuss how the analysis should be modified when party I's objective may depend on the state s as well as a, focusing on the specification $u(a, s) = -(a - b(s))^2$ where $b(s)$ -assumed to be strictly increasing- represents the action a most preferred by party I in state s.

Communication game.

In standard communication games à la Crawford and Sobel (1982), party I sends a message m once to party U who then chooses an action a. Message m need not have any accepted meaning in that approach. That is, the message space M need not be related to the state space S.

I consider the following modifications. First, in order to identify messages as lies or truths, I explicitly let all the states $s \in S$ be possible messages, that is $S \subseteq M$. When message $m = s$ is sent, it can be thought of as party I saying "The state is s.". I also allow party I to send messages outside S such as "I do not know the state" when everybody knows that I knows s, that is $M \setminus S \neq \varnothing$. While the set M will be assumed to be finite, in applications the set M is likely to be much larger than S.

Second, in order to let memory play a role, I assume that party I sends two messages

m_1, $m_2 \in M$ one after the other, at times $t = 1$ and 2. Party U observes the messages m_1, m_2, and she chooses her action $a(m_1, m_2)$ as a function of these.

Letting party I send two messages instead of one would make no difference if after sending message m_1, party I always remembered what message m_1 he previously sent, and if both parties I and U were fully rational, as usually assumed. While party U will be assumed to be rational, I consider environments in which party I at time $t = 2$ has imperfect memory about the message m_1 sent at time $t = 1$. More precisely, I assume that when party I in state s tells the (whole) truth at time $t = 1$, i.e. sends $m_1 = s$, he remembers that $m_1 = s$ at $t = 2$, but when he lies (identified here with not telling the whole truth) and sends $m_1 \neq s$, he does not remember what message m_1 he previously sent (he may still think that he sent $m_1 = s$, as I do not impose in the basic approach that he is aware that he lied, see below for further discussion).

A key modeling choice concerns how party I at time $t = 2$ forms his expectation about the message sent at $t = 1$ when he lied lie at $t = 1$. I adopt the following approach.

Solution concept

A multi-self approach is considered, which is standard in situations with imperfect recall (see Piccione and Rubinstein (1997)). That is, think of the state s as a type for party I, and envision party I with type s at times $t = 1$ and 2 as two different players $I_1(s)$ and $I_2(s)$ having the same preference as party I. To model the belief of a forgetful liar, let $\sigma_1(m \mid s)$ denote the (equilibrium) probability with which message $m_1 = m$ is sent at $t = 1$ by party I with type s. Assuming that at least one type s lies with positive probability at $t = 1$, i.e. $\sigma_1(m \mid s) > 0$ for at least one (m, s) with $m \neq s$, one can define the distribution of lies at $t = 1$ aggregating lies over all possible realizations of s. The probability of message m in this aggregate distribution is

$$\sum_{s \in S,\ s \neq m} \sigma_1(m \mid s)p(s) / \sum_{(m',s') \in M \times S,\ m' \neq s'} \sigma_1(m' \mid s')p(s'). \tag{1}$$

In an equilibrium with forgetful liars σ, when $I_1(s)$ lies at $t = 1$ (i.e. sends $m_1 \neq s$), player $I_2(s)$ at time $t = 2$ believes that player $I_1(s)$ sent m with probability as expressed in (1). If no lie is ever sent at time $t = 1$ in equilibrium, the belief after a lie can be arbitrary. By contrast, when $I_1(s)$ tells the truth (i. e., sends $m_1 = s$), player $I_2(s)$ knows that $m_1 = s$.

The other features of the equilibrium with forgetful liars are standard. All expectations other than that of $I_2(s)$ about m_1 after a lie at $t = 1$ are correct, and all players are requested to choose best-responses to their beliefs given their preferences (deviations of I are local and not joint between $t = 1$ and 2 due to the multiself specification).

To give a concrete illustration of how the beliefs about m_1 are formed by party I at $t = 2$, assume there are four states s_1, s_2, s_3 and s_4, with states s_2 and s_3 being equally likely (the other two states may have different ex ante probabilities). Assume that party I at time $t = 1$, sends message $m_1 = s_1$ in states s_1 and s_2, and message $m_1 = s_4$ in states $s = s_3$ and s_4. At $t = 2$, in states $s = s_1$ and s_4, party I remembers that he sent $m_1 = s$ as there is no lie in these states. By contrast, in states $s = s_2$ and s_3, there is a lie so that party I does not remember at $t = 2$ what message he previously sent. In these two states, party I believes that he sent as first message s_1 and s_4 with equal probability where the weighting of the two lies is imposed by the assumed time $t = 1$ communication strategy and the assumption that s_2 and s_3 are equally likely.

As is common in many studies of communication games (see for example Chen (2011) or Hart et al. (2017)), I consider refinements/perturbations which I view as natural and serve the purpose of ruling out implausible equilibria and/or ensuring the existence of pure strategy equilibria.

Refinements/Perturbations.

First, I assume that in case of indifference between lying and truth-telling, party I opts for truth-telling. Formally, for some positive ε assumed to be sufficiently small, party I's payoff as a function of (s, a, m_1, m_2) is:

$$U_I(s, a, m_1, m_2) = u(a, s) - \varepsilon(1_{m_1 \neq s} + 1_{m_2 \neq s}). \tag{2}$$

That is, a lie whether at $t = 1$ or 2 is assumed to inflict an extra ε cost. This will be referred to as the TP-perturbation.

Second, I assume that were party U to receive twice the same message $m_1 = m_2 = s$ corresponding to a state $s \in S$ that would never been sent in equilibrium, party U would make the inference that the state is s and choose $a(s, s) = s$ accordingly. This can be rationalized if there is a chance (assumed to be small) that party I is a truth-teller no

matter what the state s is, which is consistent with a number of experimental findings (see, for example, Gneezy (2005)). This will be referred to as the TB-perturbation.

Third, in order to support pure strategy equilibria, I will be assuming that if m_1 and m_2 with $(m_1, m_2) \notin S^2$ are received while (m_1, m_2) is not supposed to be sent in equilibrium, then party U makes the inference that s is low enough, and to simplify the exposition of the analysis, I will let $a(m_1, m_2) = 0$ in such a case.[10]

Finally, I will be making the following assumption where for any subset T of S, $p(T)$ denotes $\Pr(s \in T)$ and $e(T)$ denotes $E(s \mid s \in T)$.[11]

Genericity assumption (GE). For any families $(T_a^k)_k$, $(T_b^k)_k$, and $(T_c^k)_k$, of disjoint (non-empty) subsets of S, if

$$\sum_k p(T_a^k)(e(T_b^k) - e(T_c^k)) = 0$$

then $T_b^k = T_c^k$ for all k.

In the next Sections, I characterize the equilibria with forgetful liars of the above communication game assuming that ε is small enough, i.e., smaller than half the minimum value of $\sum_k p(T_a^k)(e(T_b^k) - e(T_c^k))$ when allowing T_a and $T_b \neq T_c$ to vary, which by the genericity assumption is strictly positive. The interpretation and use of assumption GE will be as follows. At $t = 2$ after a lie at $t = 1$, party I will have the belief in a pure strategy equilibrium that message $m_1 = m^k$ was sent at $t = 1$ with probability $p(T_a^k)$. For each such message m^k, party I will expect that if $m_2 = m^b$ or m^c is sent at $t = 2$, the action $a_b^k = e(T_b^k)$ or $e(T_c^k)$ will be observed next. Assumption GE will then guarantee that party I finds only one message strictly optimal at $t = 2$ in a pure strategy equilibrium, no matter what $(T_a^k)_k$, $(T_b^k)_k$, $(T_c^k)_k$ are.

[10] An alternative would be to let $a(m_1, m_2)$ be a free variable in this case and characterize the corresponding set of all such equilibria. It turns out that either approach would yield similar insights, thereby leading me to adopt the simpler one.

[11] Observe that it holds generically whenever there are at least three states where genericity can either be defined over the values of $s_2, ..s_{n-1}$ letting the probabilities $p(s_k)$ fixed or over the probabilities $p(s_1), ...p(s_n)$ letting the states s_k fixed.

Comments.

1. The chosen modeling of a liar's expectation assumes that to form his expectation about his time $t = 1$ lie, party I considers the overall distribution of lies as observed in similar interactions (played by other economic agents) aggregating over all possible realizations of s. The equilibria with forgetful liars as defined above correspond to steady states of such environments. Given the aggregation over states, this approach can be embedded in the general framework of the analogy-based expectation equilibrium (Jehiel (2005) and Jehiel and Koessler (2008)).[12] It should be mentioned that the underlying learning environment supporting the approach requires that the messages be tagged as lies before the state is disclosed so that the aggregate distribution of lies is accessible, but the joint distribution of messages and states is not.[13] If instead, messages are simply disclosed with no mention whether they are lies, a possible alternative specification is that the belief of a liar would match the aggregate distribution of time $t = 1$ messages. I will briefly discuss the implications of such an alternative specification later.

It should be highlighted that in the above interpretation, party I when lying at $t = 1$ should be viewed at time $t = 2$ as not remembering his time $t = 1$ strategy, which I motivated on the ground that each individual in the role of party I is an occasional player. If instead the forgetful liar remembers his strategy, the knowledge of the state s together with the strategy would lead party I to have a different belief. More precisely, in state s, party I at time $t = 2$ after party I lied at $t = 1$ should expect that m was sent at $t = 1$ with probability

$$\sigma_1(m \mid s) / \sum_{m' \in M,\, m' \neq s} \sigma_1(m' \mid s) \tag{3}$$

[12] Formally, define the interaction using the following extensive form game. Nature chooses the state s. Player $I_1(s)$ who observes s decides first whether to tell the truth (Truth) or lie (Lie). If he chooses Truth, he sends $m_1 = s$. If he chooses Lie, he chooses $m_1 \in M$ (with the requirement that $m_1 \neq s$). Then player $I_2(s)$ who observes s chooses m_2 observing from past play whether player $I_1(s)$ chose Truth or Lie. Finally, party U chooses a with payoffs as defined in the main text. The analogy partition considered by player I_2 puts all the m_1 decision nodes of players $I_1(s)$ following the Lie choice into one analogy class. All other decision nodes are singleton analogy classes. It is readily verified that the equilibria with forgetful liars as defined in the main text are analogy-based expectation equilibria of the proposed strategic environment.

[13] In many contexts such as the criminial investigation one, I would argue such a disclosure scenario is natural as evidence of lies together with the associated messages are typically disclosed much before the details of the state are disclosed (which are typically disclosed after many additional investigations).

whenever $\sum_{m' \in M,\ m' \neq s} \sigma_1(m' \mid s) > 0$. In other words, party I when lying at $t = 1$ would form his expectation about his lie by conditioning the equilibrium distribution of lies on the state s (that he is assumed to remember).[14] This approach is in the spirit of either of the multiselves approaches to imperfect recall as defined in Piccione and Rubinstein (1997). While the main analysis is developed with the expectation formulation (1), I will also mention the implications of the expectation formulation (3) in pure persuasion situations.

2. In the approach developed above, I assume that player $I_2(s)$ when a lie was made by $I_1(s)$ is not aware that $I_1(s)$ lied and accordingly can assign positive probability to $m_1 = s$ in his belief as defined in (1) if it turns out that $m_1 = s$ is a lie made with positive probability by some type $s' \neq s$.[15] If instead such a player $I_2(s)$ were aware he made a lie, it would then be natural for him to rule out that $m_1 = s$, and a new definition of belief (conditioning the aggregate distribution of lies on $m_1 \neq s$) should be considered.[16] I will discuss the implications of such an alternative specification later.

3 Pure persuasion

In this Section, I assume that for all a and s, $u(a, s) = a$. Thus, whatever the state s, party I wants the belief held by party U about the expected value of s to be as high as possible.

[14] Another possible interpretation of expectation (3) assuming that economic agents play only once is that party I would have access from past plays to the joint distribution of lies and states, which would allow him to construct the conditional distributions. In many cases of interest though, the joint distribution is not so clearly accessible, making this interpretation less natural.

[15] For example, in the context of the concrete illustration provided after defining expectation (1), if at $s = s_2$, party I sends $m_1 = s_3$ instead of $m_1 = s_1$, in state $s = s_3$, player $I_2(s)$ believes that he either sent $m_1 = s_3$ or s_4 with equal probability.

[16] That is, the belief a liar I in state s^* should be replaced by

$$\sum_{s \in S \setminus \{s^*\},\ s \neq m} \sigma_1(m \mid s) p(s) / \sum_{(m', s') \in M \times S \setminus \{s^*\},\ m' \neq s'} \sigma_1(m' \mid s') p(s').$$

3.1 Pure strategy equilibria

A simple class of strategies.

As will be shown below, equilibria with forgetful liars employing pure strategies will involve a simple class of communication strategies that I now describe. For any $(s^l, s^h) \in S^2$ with $s^l < s^h$, the (s^l, s^h)-communication strategy is defined as follows. Party I in state $s \in S$ sends twice the same message $m_1(s) = m_2(s)$. Party I with type $s \leq s^l$ lies and tells s^h , i.e. $m_1(s) = m_2(s) = s^h$, and party I with type $s > s^l$ tells twice the truth, i.e. $m_1(s) = m_2(s) = s$.

Several simple observations follow whenever party I employs the (s^l, s^h) communication strategy. First, there are no inconsistent messages $m_1 \neq m_2$ being sent in equilibrium. Second, the induced aggregate distribution of lie at $t = 1$ is a mass point on s^h.

Third, the best-response of party U is to choose $a(s, s) = s$ for $s \in S$ whenever $m_1 = m_2 = s \neq s^h$ and $a(s^h, s^h) = a^E(s^l, s^h) = E(s \mid s \leq s^l \text{ or } s = s^h)$ when $m_1 = m_2 = s^h$.

Fourth, if party I with type $s \leq s^l$ were to tell the truth $m_1 = m_2 = s$, he would induce action $a = s$ instead of $a^E(s^l, s^h)$. So a necessary condition for the (s^l, s^h)-communication strategy to be part of an equilibrium is that (s^l, s^h) satisfies $a^E(s^l, s^h) - 2\varepsilon \geq s^l$.

Fifth, if party I with type $s > s^l$ were to lie at time $t = 1$, he would believe at time $t = 2$ that he sent $m_1 = s^h$ according to the proposed solution concept. By lying and sending $m_1 = s^h$ at time $t = 1$, party I with type s could ensure to get $a^E(s^l, s^h) - 2\varepsilon$, since after a lie at $t = 1$, player $I_2(s)$ would find it optimal to send s^h expecting to get $a^E(s^l, s^h) - 2\varepsilon$ rather than any other message m_2 that would give him at most $-\varepsilon$. Thus, letting $s^l_+ = \min\{s_k \neq s^h \text{ such that } s_k > s^l\}$, another necessary condition for the (s^l, s^h)-communication strategy to be part of an equilibrium is that $s^l_+ \geq a^E(s^l, s^h) - 2\varepsilon$.

Focusing on the communication strategy of party I, equilibria with forgetful liars that employ pure strategies are characterized as follows.

Proposition 1 *An equilibrium with forgetful liars in pure strategies always exists. It either takes the form that no lie is being made or it takes the form that party I uses an (s^l, s^h)-communication strategy for some (s^l, s^h) satisfying $s^l_+ \geq a^E(s^l, s^h) - 2\varepsilon \geq s^l$. Any (s^l, s^h)-communication strategy satisfying the latter requirements can be part of an equilibrium with forgetful liars.*

That an equilibrium with truth-telling in all states s can be sustained is easily established by letting player $I_2(s)$ believe after a lie of player $I_1(s)$ that player $I_1(s)$ sent $m_1 = 0$ with probability 1 (and letting $a(s,s) = s$ for all $s \in S$, as implied by the truth-telling equilibrium behavior of party I). That truth-telling can be an equilibrium clearly illustrates the deep effects of the forgetfulness of liars on the strategic analysis, as without memory problems no full revelation of the state could be expected in any Nash equilibrium. I note however that the truth-telling equilibrium is somehow fragile, as it requires for vanishingly small lying costs that in case of lie, the belief is that only $m_1 = 0$ was sent at $t = 1$. Given that the purpose of lies is to make party U believe that the state s is as high as possible, it seems odd to have that all lies are expected to be concentrated on 0, when these do not arise in equilibrium, thereby making the truth-telling equilibrium not very plausible in my view.[17]

It is then of interest to understand the properties of the equilibria with forgetful liars other than truth-telling, which focusing on equilibria that employ pure strategies is covered by Proposition 1. Consider a pure strategy equilibrium in which there is some lying activity, that is, at least one type lies either at $t = 1$ or 2 . Refer to $\sigma_1(s)$ and $\sigma_2(s)$ as the corresponding messages sent in state s by $I_1(s)$ and $I_2(s)$, respectively. I decompose the characterization of such equilibria in a few elementary steps, the combination of which will establish Proposition 1. Missing proofs appear in Appendix.

A preliminary observation is that in such an equilibrium, lies cannot only take place at $t = 2$. That is, there must be some lies at $t = 1$. To see this, observe by contradiction that if there were no lie at $t = 1$, and in state s, player $I_2(s)$ were to lie at $t = 2$, the message profile $(m_1 = s, \sigma_2(s))$ would perfectly reveal the state s to party U. As a result, player $I_2(s)$ would be strictly better off telling the truth rather than $\sigma_2(s) \neq s$, as this would allow player $I_2(s)$ to save the lying cost associated to $\sigma_2(s) \neq s$.[18] This observation implies that there is a well defined on-the path aggregate distribution of lies at $t = 1$ in any pure strategy equilibrium in which there is some lying activity.

[17] To see that truth-telling is no longer an equilibrium when lies are not expected to be concentrated on 0, consider party I in state $s = 0$. After a lie of $I_1(s = 0)$, player $I_2(s = 0)$ would strictly prefer sending any message $m_2 \in S \setminus \{0\}$ belonging to the expected support of lies rather than $s = 0$, at least for small enough lying costs ε. Anticipating the most preferred lie of $I_2(0)$, player $I_1(0)$ could then make the same lie, thereby leading to a profitable deviation.

[18] It would lead to the same action $a = s$ as $(s, \sigma_2(s))$ given that $m_1 = m_2 = s$ would be off-the-path.

The next observation is that whenever party I lies at $t = 1$, there is a unique message m^* that can be sent at $t = 2$. Formally,

Lemma 1 *In any pure strategy equilibrium, there exists a message $m^* \in M$ such that whatever s, if $\sigma_1(s) \neq s$ then $\sigma_2(s) = m^*$.*

The logic of this result is that after a lie at $t = 1$, player $I_2(s)$ has the same belief about the message sent at $t = 1$, irrespective of s. Given that up to the lying costs, the preferences of $I_2(s)$ are the same whatever s, the genericity assumption (GE) together with the assumption that ε is small enough guarantees that, after a lie at $t = 1$, player $I_2(s)$ finds a unique message to be optimal, no matter what s is.[19]

Given m^* as introduced in Lemma 1, define

$$L(m^*) = \{s \in S \text{ such that } \sigma_1(s) = \sigma_2(s) = m^*\}$$

as the set of states s such that in equilibrium the same message m^* is consistently sent at $t = 1$ and 2.

Similarly, for any $m \neq m^*$, define

$$S_{inc}(m, m^*) = \{s \in S \text{ such that } \sigma_1(s) = m \text{ and } \sigma_2(s) = m^*\}$$

as the set of states s such that in equilibrium message $m_1 = m$ is sent at $t = 1$ and message $m_2 = m^*$ is sent at $t = 2$.

The following observation is derived from an argument similar to the unravelling argument used in certification games.

Lemma 2 *One must have $m^* = \max\{s \in L(m^*)\}$.*

Roughly, Lemma 2 holds because otherwise party I in state $s^* = \max\{s \in L(m^*)\}$ would prefer telling the truth (that would uniquely identify the state) rather than lying and sending m^*.

[19] Specifically, at $t = 2$ after a lie at $t = 1$, party I must have the belief in a pure strategy equilibrium that message $m_1 = m^k$ was sent at $t = 1$ with some probability $p(T_a^k) = \Pr(s \in T_a^k)$ for some disjoint subsets T_a^k of S. For each such message m^k, party I will expect that if $m_2 = m^b$ or m^c is sent at $t = 2$, the action $a_b^k = e(T_b^k)$ or $e(T_c^k)$ will be observed next -where $e(T) = E(s \mid s \in T)$ and T_b^k (resp. T_c^k) is the subset of states $s \in S$ such that (m^k, m^b) (resp. (m^k, m^c)) is sent in equilibrium. The genericity assumption ensures that there is a unique best-response whatever the sets T_a^k, T_b^k, T_c^k are.

The next observation informs us about the possible time $t = 2$ messages after player $I_1(s)$ has told the truth at $t = 1$. Such a message cannot be a lie other than m^* as the corresponding message profile would reveal the state and make a deviation to truth-telling profitable.

Lemma 3 *If in equilibrium player $I_1(s)$ tells the truth, i.e. $\sigma_1(s) = s$, then either $\sigma_2(s) = s$ or m^*.*

Lemmas 1 and 3 show that if player $I_2(s)$ lies, one should have $\sigma_2(s) = m^*$ whether player $I_1(s)$ lies (Lemma 1) or tells the truth (Lemma 3). They also establish that if the messages m_1 and m_2 are not the same, one must have $m_2 = m^*$. The next lemma establishes using again an unravelling argument that there cannot be inconsistent messages in a pure strategy equilibrium. Formally,[20]

Lemma 4 *For any $m \neq m^*$, one must have $S_{inc}(m, m^*) = \varnothing$.*

The final steps of the proof of Proposition 1 are as follows. By the above lemmas, the only possible lie at $t = 1$ is m^* with $m^* = \max\{s \in L(m^*)\}$, and there can be no inconsistent messages m_1, $m_2 \neq m_1$ being sent in equilibrium. This implies that the aggregate distribution of time $t = 1$ lies is a mass point concentrated on m^*. If $m^* = 0$, these conditions imply that we are in a truthful equilibrium, and thus to the extent that there are some lies being made, $m^* = 0$ can be ruled out.

It is then easily seen that after a lie at $t = 1$, irrespective of the state s, player $I_2(s)$ strictly prefers sending m^* expecting to get no less than $a(m^*, m^*) - 2\varepsilon$ rather than any other message expecting to get at most $-\varepsilon$ (as such a message profile would be off-the-path and results in action $a = 0$). And if player $I_1(s)$ tells the truth, it is optimal for player $I_2(s)$ to tell the truth as well, since any other message would result in action $a = 0$ and induces an extra lying cost.

Thus, party I in state $s \neq m^*$ either lies twice by sending $m_1 = m_2 = m^*$ expecting to get $a(m^*, m^*) - 2\varepsilon$ or he tells the truth twice $m_1 = m_2 = s$ expecting to get $a(s) = s$ (given than $m_1 = m_2 = s$ can safely be attributed to state s by party U, since $s \neq m^*$ would not be a lie made by party I in equilibrium no matter what the state is).

[20] Lemma 4 is shown by observing that if $S_{inc}(m, m^*) \neq \varnothing$ then party I in state $s^*_{inc} = \max S_{inc}(m, m^*)$ would strictly gain by telling the truth.

As a result, party I in state $s \neq m^*$ chooses to lie by sending twice m^* whenever $s < a(m^*, m^*) - 2\varepsilon$, and he tells the truth twice whenever $s > a(m^*, m^*) - 2\varepsilon$. In state $s = m^*$, party I tells the truth, but his message also happens to be the common lie made in equilibrium. In turn, this implies that in a pure strategy equilibrium, $a(m^*, m^*)$ takes the value $a^E(s^l, s^h)$ as defined above with $s^h = m^*$ and s^l being such that $s^l_+ \geq a^E(s^l, s^h) - 2\varepsilon \geq s^l$. Proposition 1 is thus established.

Several additional remarks follow. First, I observe that an equilibrium in pure strategy with some lying activity always exists. This is shown by letting $s^l = 0$, $s^h = s_2$ and observing then that $s^l_+ > s_2 \geq a^E(s^l, s^h) - 2\varepsilon \geq s^l$. More generally, the pure strategy equilibria other than truth-telling are parameterized by the common lie $m^* \in S \setminus \{0\}$, and for every such m^* one can show that there is an equilibrium with forgetful liars employing pure strategies in which for some (s^l, s^h) with $s^h = m^*$ party I follows the (s^l, s^h)- communication strategy.[21]

Second, I note some close analogy between the shape of a pure strategy equilibrium with lie $m^* \in S \setminus \{0\}$ and the Perfect Bayesian Nash equilibria that would arise in the one-round communication game in which party I with type $s \in S$ could certify his type when $s \neq m^*$ but not when $s = m^*$.[22] Even though there is no explicit certification technology in my setting, if party I lies at $t = 1$ in a pure strategy equilibrium with lie m^*, he anticipates that he will be sending m^* at $t = 2$. Thus, in such an equilibrium, the choice for party I with type $s \neq m^*$ boils down either to be telling the truth at $t = 1$ and 2, resulting in outcome $a(s) = s$ -the same outcome as the one party I would obtain if his type s were disclosed-, or consistently sending the lie $m_1 = m_2 = m^*$ (that results in action $a(m^*)$, which is endogenously determined as in the certification framework).

Finally, while I regard the perturbations/refinements introduced at the end of Section 2 as fairly reasonable ones in the context of communication games, it is instructive to know how the analysis would be affected when alternative specifications for off-path actions are considered, when perturbations TP or TB are removed, or when the genericity assumption

[21] Observe that for all $s^h \in S \setminus \{0\}$ we have that $a^E(0, s^h) > 2\varepsilon$, and $a^E(s^h, s^h) - s^h < 0$. Thus, choosing s^l to be $\max_s\{s \text{ such } a^E(s, s^h) > s + 2\varepsilon\}$ guarantees that all conditions are satisfied.

[22] With states varying on the continuum, a similar situation has first been considered by Dye (1985) who extended the classic persuasion models analyzeded by Grossman (1981) and Milgrom (1981) by adding the possibility that party I would be uninformed and would be unable to prove (or certify) that he is uninformed. Somehow type m^* plays a role similar to the uninformed type in Dye.

GE is dropped. I now provide a sketchy description of this.

1) The above analysis of pure strategy equilibria remains unchanged for alternative specifications of $a(m_1, m_2)$ for off-the-path message profiles (m_1, m_2), as long as these are set to be small enough. If by contrast such $a(m_1, m_2)$ are set too big, then there is no pure strategy equilibrium, and one has to look for mixed strategy equilibria. Thus, the assumption that $a(m_1, m_2) = 0$ for off-the-path message profiles is without loss of generality for the analysis of pure strategy equilibria.[23]

2) If the genericity assumption GE is dropped, more equilibrium outcomes possibly with multiple lies and different welfare consequences can sometimes be supported. For example, suppose one considers a situation with an even number n states such that $s_{n-k+1} + s_k = 1$ for all k and an equal probability that each state arises. One can support an equilibrium with forgetful liars in which for each $s \in \{s_k, s_{n+1-k}\}$ party I consistently reports that the state is $\max\{s_k, s_{n+1-k}\}$ at $t = 1, 2$. In this case, the aggregate distribution of lie is the uniform distribution over $\max\{s_k, s_{n+1-k}\}$ for the various k. After any such (consistent) lie, party U would choose $a = \frac{1}{2}$, and party I after a lie at $t = 1$ would be indifferent as to which lie in the set $\{\max\{s_k, s_{n+1-k}\}\}_k$ to choose. By requiring that in state $\min\{s_k, s_{n+1-k}\}$ the lying party I chooses $\max\{s_k, s_{n+1-k}\}$ at $t = 2$, one can support this as an equilibrium. Note though the fragility of such a construction, as it would require that, in state $s = \min\{s_k, s_{n+1-k}\}$, party I at $t = 2$ chooses the specific best-response $\max\{s_k, s_{n+1-k}\}$ when in fact he is indifferent between all $\max\{s_{k'}, s_{n+1-k'}\}$ obtained when k' varies.[24,25]

3) If perturbation TB concerning the interpretation of $m_1 = m_2 = s_k$ off-the path is removed, then one can support equilibria in which party I lies in more states. In particular,

[23] To ensure that there are no other pure strategy equilibria for generic values of the states, one should assume that the perturbations giving rise to the choices of such $a(m_1, m_2)$ are not fine-tuned to the values of s^k, as would result from trembling behaviors assumed to be solely determined by the order of the states (and not their exact values).

To ensure that the truth-telling trembling dominates the alternative trembling possibly resulting in inconsistencies (so that $a(s, s) = s$ whenever $m_1 = m_2 = s \in S$ is off-the-path), one should have in mind that the message space is much larger than the state space so that being consistently truthful by chance (i.e., without being a truth-teller) would be very unlikely.

[24] If party I were choosing another best-response at $t = 2$, he would be caught sending inconsistent messages, and party I would rather avoid sending message $\max\{s_k, s_{n+1-k}\}$ at $t = 1$.

[25] This argument illustrates why an alternative to the GE assumption to get the same result as in Proposition 1 is to assume that in case of indifferences, (the lying) party I always picks the same best-response irrespective of the state.

party I consistently lying and sending $m_1 = m_2 = 1$ in all states $s \neq 1$ can be part of an equilibrium with forgetful liars if the expectation is that when $m_1 = m_2 = s_k$ with $s_k \neq 1$ are received a sufficiently low action (for example, $a = 0$) would be chosen by party U. To see this, observe that with the proposed strategy, the action after $m_1 = m_2 = 1$ would just be the expected value $E(s_k)$ of the state, there would only be one lie $m^* = 1$ in equilibrium, and telling the truth consistently at $t = 1$ and 2 would not be attractive to party I when the state is $s_k \neq 1$. More technically, when perturbation TB is removed, the unravelling argument breaks down, and party I in states s that are different from the common lie but above the action resulting from the common lie may prefer lying to telling the truth. The breakdown of the unravelling argument in turn allows to support more equilibrium outcomes with different welfare consequences, thereby revealing the essential role of perturbation TB in the derivation of Proposition 1.

4) If perturbation TP concerning the slight preference for truth-telling is removed, the insight that inconsistent messages cannot arise in equilibria employing pure strategies (lemma 4) no longer holds. That is, new equilibria in which party I in states $s > a(m^*, m^*)$ with $s \neq m^*$ would be sending a message $m_2 \neq s$ after $m_1 = s$ was sent can now be supported (this is so because the inconsistency (s, m_2) would safely be attributed to state s, thereby leading to action $a = s$ in this case). I note that such equilibria (which are outcome equivalent to those considered in Proposition 1) would not be robust to other (natural) perturbations in which party I in sufficiently low states would be viewed as randomly sending inconsistent messages with positive probability (since with such additional perturbations, (s, m_2) would now be followed by a lower action than when (s, s) is chosen).[26]

3.2 Approximate first-best with fine grid

So far, states s_k could be distributed arbitrarily on $[0, 1]$. What about the case when consecutive states are close to each other and all states have a comparable ex ante probability? I show that in such a case, all equilibria in pure strategies are close to the truth-telling equilibrium, resulting in the approximate first-best outcome for party U. More precisely,

[26] Based on this, I would argue that perturbation TP may not be needed for the derivation of Proposition 1 if other (plausible) perturbations are considered instead.

Definition 1 *A state space $S^n = \{s_1, ...s_n\}$ satisfies the n-fine grid property if $s_{k+1} - s_k < \frac{2}{n}$ for all k, and for some $(\underline{\alpha}, \overline{\alpha})$, $0 < \underline{\alpha} < \overline{\alpha}$, set independently of n, $\underline{\alpha} < p(s_k)/p(s_{k'}) < \overline{\alpha}$, for all k, k'.*

I will be considering sequences of state spaces S^n satisfying the n-fine grid assumption and of lying costs ε^n where for each n (the above genericity assumption (GE) is satisfied and) ε^n is smaller than half the minimum value of $\sum_k p(T_a^{k,n})(e(T_b^{k,n}) - e(T_c^{k,n}))$ when allowing $T_a^n = (T_a^{k,n})_k$ and $T_b^n = (T_b^{k,n})_k \neq (T_c^{k,n})_k = T_c^n$ to be any families of disjoint subsets of S^n.

Proposition 2 *Consider a sequence $(S^n, \varepsilon^n)_{n=\underline{n}}^{\infty}$ satisfying the above conditions, and a sequence $(\sigma^n)_{n=\underline{n}}^{\infty}$ of pure strategy equilibria with forgetful liars associated with (S^n, ε^n). For any $\widehat{a} > 0$, there exists $\overline{n}$ such that for all $n > \overline{n}$, the equilibrium action of party U after a lie prescribed by σ^n is smaller than $\widehat{a}$. As n approaches ∞, the expected utility of party U approaches the first-best (i.e. converges to 0).*

To prove Proposition 2, I make use of the characterization result of Proposition 1. Let a^* be the expected payoff obtained by party I when lying at $t = 1$ in an equilibrium in pure strategy. If a^* is significantly away from 0, say bigger than $\widehat{a}$ assumed to be strictly positive, then under the fine grid property the expectation of s over the set of states that are either below a^* or else equal to 1 must be significantly below a^*. But then $I(s)$ for some $s = s_k$ strictly below a^* would strictly prefer telling the truth rather than lying undermining the construction of the equilibrium (that requires party $I(s)$ with $s < a^*$ to be lying). This argument shows that a^* must get close to 0 as n approaches ∞, thereby paving the way to prove Proposition 2.

The intuition for Proposition 2 can be understood as follows. For a given lie m^* to possibly emerge in equilibrium, it should be that the probability that the state $s = m^*$ arises is not too small relative to the probability that the lie m^* is used. In the fine grid case, this implies that there can be little lying in such an equilibrium, since the probability of each state becomes increasingly small in this case. Another way to think of this result is to build on the observation made after Proposition 1. An equilibrium with forgetful liars in pure strategy with lie m^* can be viewed as a Perfect Bayesian Nash equilibrium

of a certification game in which party I can certify his type when $s \neq m^*$, but not when $s = m^*$. In such a certification framework, if the ex ante probability of m^* gets small -which must be so in the fine grid case- one gets an equilibrium outcome close to that in the classic persuasion game in which the unravelling argument leads to full disclosure (and no lying).

3.3 Mixed strategy equilibria

I now consider mixed strategy equilibria. I will not aim at characterizing all such equilibria, but instead I will consider a subclass of those having the property that, irrespective of s, when party I lies, he randomizes, and, for some $(m_k^*)_k$ and some $(\mu_k)_k$, chooses message m_k^* with probability μ_k in the same way and independently at $t = 1$ and 2. It should be noted that such a restriction would arise in all mixed strategy equilibria, if I were to assume that in case of indifference, the chosen randomization over messages is not allowed to depend on the state s nor on the calendar time t (see the discussion paper version Jehiel (2019) in which such a feature is imposed as a refinement).

Consider such a mixed strategy equilibrium that necessarily involves multiple lies. I note that some inconsistencies must arise with positive probability on-the-path, and any inconsistent messages (m_1, m_2) with $m_1 \neq m_2$ arising on-the-path must result in the same action of party U denoted hereafter a_{inc}, since any such message profile would be equally informative about the state s. Moreover, the optimality condition for liars would impose, letting $a_k = a(m_k^*, m_k^*)$, that $\mu_k a_k + (1 - \mu_k)a_{inc}$ is independent of k. This common value will be denoted a^* hereafter.

I next observe that all m_k^* must belong to S (as results from an unravelling argument) and that party I in state m_k^* should be telling the truth twice (given that some types must find the lie m_k^* weakly optimal, it must be that in state m_k^*, party I strictly prefers telling the truth rather than lying that would impose an extra lying cost).

Moreover, take any $s \in S$ other than m_k^* for $k = 1, ..K$. If $s < a^* - 2\varepsilon$, $I_1(s)$ would strictly prefer sending any m_k^* expecting to get $a^* - 2\varepsilon$ rather than telling the truth that would only yield s. If $s > a^* - 2\varepsilon$, $I_1(s)$ would strictly prefer telling the truth (anticipating that $I_2(s)$ would also do so) rather than lying. These observations yield.

Proposition 3 *The following define a class of mixed strategy equilibria. For some* a^*, m_k^*, $k = 1...K$, *with* $m_k > a^*$, *and* $\mu_k > 0$ *with* $\sum_k \mu_k = 1$, *satisfying*
$\mu_k a_k + (1 - \mu_k) a_{inc} = a^*$
$a_{inc} = E(s \in S, s < a^* - 2\varepsilon)$
$a(m_k^*, m_k^*) = a_k$ *with*
$a_k = (\mu_k \Pr(s \in S, s < a^* - 2\varepsilon) a_{inc} + p(m_k^*) m_k^*) \,/\, (\mu_k \Pr(s \in S, s < a^* - 2\varepsilon) a_{inc} + p(m_k^*) m_k^*)$
and
$a(s, s) = s$ *for* $s \in S$, $s \neq m_k^*$, $k = 1, ...K$,

$I_t(s)$ *with* $s < a^* - 2\varepsilon$ *sends* m_k^* *with probability* μ_k *independently at* $t = 1, 2$
$I_t(s)$ *with* $s > a^* - 2\varepsilon$ *tells the truth at* $t = 1, 2$.

Observe that $a_k > a_{inc}$ for all k, and thus being inconsistent is never profitable relative to what happens when the same message is being sent at $t = 1$ and 2 on-the-path. Such a finding can be viewed as formalizing that being inconsistent in a strategic communication setting with forgetful liars must be harmful. Observe also that as for the pure strategy equilibria, in the fine grid case, the proposed mixed strategy equilibria approach the first-best for party U, as can be inferred from the observation that a^* must be converging to 0 in such a limit (see the working paper version for details on this).

3.4 On alternative modeling of forgetful liars

3.4.1 When the informed party knows his lying strategy

How is the analysis affected when considering the scenario in which a forgetful liar would know the distribution of lies conditional on the state (and not just in aggregate over the various states as assumed above, see expression (3)).

While the equilibria arising with the main proposed approach would continue to be equilibria with this alternative approach, the main observation is that many additional equilibrium outcomes can also arise. In particular, even in the fine grid case, equilibrium outcomes significantly away from the first-best can now be supported. To illustrate this, I focus on equilibria employing pure strategies. Consider a setup with an even number n of states and a pairing of states according to $S_k = \{\underline{s}_k, \overline{s}_k\}$ with $(S_k)_k$ being a partition of the state space and $\underline{s}_k < \overline{s}_k$ for all k. I claim that with this alternative approach, one can

support an equilibrium in which for every k, $I(\underline{s}_k)$ lies consistently by sending $m_t = \overline{s}_k$ at $t = 1, 2$ while $I(\overline{s}_k)$ tells the truth. To complete the description of the equilibrium, party U's action when hearing twice $\overline{s}_k$ should be $a(\overline{s}_k, \overline{s}_k) = E(s \in S_k)$, and I let the belief of $I_2(\overline{s}_k)$ if $I_1(\overline{s}_k)$ were to lie to be that message 0 was sent at $t = 1$.[27]

The reason why such an equilibrium can arise now is that with the new expectation formulation, when $I_1(\underline{s}_k)$ lies at $t = 1$, player $I_2(\underline{s}_k)$ (rightly) believes that player $I_1(\underline{s}_k)$ sent $m_1 = \overline{s}_k$ given that this is the only lie made by $I_1(\underline{s}_k)$ in equilibrium. As a result, player $I_2(\underline{s}_k)$ after a lie at $t = 1$ finds it optimal to send $m_2 = \overline{s}_k$ as any other message is perceived to trigger action $a = 0$, which is less than $E(s \in S_k)$. Given that $I_1(\underline{s}_k)$ has the correct expectation about $I_2(\underline{s}_k)$' strategy, $I_1(\underline{s}_k)$ either lies and sends $m_1 = \overline{s}_k$ or else he tells the truth. Given that $E(s \in S_k) > \underline{s}_k$, he strictly prefers lying (whenever ε is small enough), thereby showing the optimality of $I_t(\underline{s}_k)$' strategy for $t = 1, 2$. Showing the optimality of $I_t(\overline{s}_k)$' strategy is easily derived using the off-path beliefs proposed above.[28]

The key reason why multiple lies can be sustained now and not previously is that the belief of $I_2(\underline{s}_k)$ after a lie at $t = 1$ now depends on $\underline{s}_k$ given that the mere memory of the state $\underline{s}_k$ together with the knowledge of the equilibrium strategy of $I_1(\underline{s}_k)$ allows player $I_2(\underline{s}_k)$ to recover the lie made by $I_1(\underline{s}_k)$, even if he does not directly remember m_1.

It is also readily verified that such equilibrium outcomes can lead party U to get payoffs bounded away from the first-best, even in the fine grid case as the number of states gets large, in contrast to the insight derived in Proposition 2 (think for example, of the limit pairing of s and $1 - s$ in the approximately uniform distribution case that would result in party U choosing approximately action $a = \frac{1}{2}$ in all states, which corresponds to what happens in the absence of any communication).

Thus, when party I knows his lying strategy (possibly as a consequence of playing the game many times), party I may still withhold a lot of information, even when physically

[27] As in the main model, one also requires that when hearing hearing off-the-path message profiles, party U chooses $a = 0$.

[28] One may be willing to refine the off-path beliefs of $I_2(\overline{s}_k)$ in the above construction for example by requiring that a lie $m_1 = 1$ (instead of $m_1 = 0$) is more likely to occur when $I_1(\overline{s}_k)$ lied (and $\overline{s}_k \neq 1$). Note that the above proposed strategies would remain part of an equilibrium with this extra refinement, assuming that $\{0, 1\}$ is one of the pairs S_k and $E(s = 0 \text{ or } 1)$ takes the smallest value among all $E(s \in S_k)$ (think of assigning sufficient weight on the state being $s = 0$). Indeed, in such a scenario, if $I_1(\overline{s}_k)$ were to lie, he would send $m_1 = 1$ anticipating that $I_2(\overline{s}_k)$ would send $m_2 = 1$ next, and this would be worse than truth-telling.

forgetting his past lies. This was not so (in particular in the fine grid case) when subjects in the role of party I were viewed as occasional players and access to past interactions was focused on the distribution of lies (and not the joint distribution of lies and states).

3.4.2 When others' lies are not tagged as such

Having again in mind that subjects in the role of party I are occasional players and learning environments in which there is no access to the joint distribution of messages and states, one may in contrast to the main modeling approach consider situations in which the time $t = 1$ messages m_1 would not, for learning purposes, be tagged as lies before the state is disclosed. In this case, there would be no easy access for new comers to the aggregate distribution of lies, and it is then more natural to assume that when party I lies at $t = 1$, he believes at $t = 2$ that he sent a message according to the aggregate equilibrium distribution of messages used at $t = 1$ (aggregating this time not only over the states but also whether or not messages correspond to the truth).

I will not develop a full analysis with this alternative formulation, but it may be interesting to note that whenever the probability $p(s_n)$ of state $s_n = 1$ is no smaller than $p(s_k)$ for every $k < n$, then the (s^l, s^h)-communication strategy with $s^h = s_n = 1$ and s^l defined appropriately so that (s^l, s^h) satisfies the conditions shown in Proposition 1 would continue to be a pure strategy equilibrium in this alternative approach. Roughly, the reason why this holds true is that with such a communication strategy, there would be enough probability on the message s^h in the aggregate distribution of time $t = 1$ messages so that a liar at time $t = 2$ would always find it optimal to send message s^h.[29] It is also not difficult to see using arguments similar to the ones developed above that with this alternative approach, pure strategy equilibria will only have one lie, there would be no inconsistent messages, and party I would have to be using (s^l, s^h)-communication strategies with the restrictions imposed in Proposition 1. Possibly, not all of the communication strategies shown in Proposition 1 could arise as equilibria, as for some low enough s^h, a liar at $t = 2$

[29] This follows because letting $a^* = E(s \text{ such that } s \leq s^l \text{ or } s = s^h)$, one would have:

$$a^* = \frac{\Pr(s \leq s^l)E(s \leq s^l) + p(s_n)s_n}{\Pr(s \leq s^l) + p(s_n)}$$

and thus $(\Pr(s \leq s^l) + p(s_n))a^* > p(s_n)s_n > p(s_k)s_k$ for any $k < n_k$.

would end up preferring message $m_2 \neq s^h$, undermining the equilibrium construction. In particular, the truth-telling strategy would not longer be part of an equilibrium. Overall, the implications of this alternative approach are very similar to the ones obtained with the main model, as far as pure strategy equilibria are concerned.

3.4.3 When liars remember that they lied

In the main approach, I assumed that a forgetful liar in state s could consider that he previously sent s with some positive probability if s happened to be a lie made in another state s'. If instead player $I_2(s)$ after a lie at $t = 1$ were to be aware that party I lied at $t = 1$, it would be more natural to assume that player $I_2(s)$ would rule out that player $I_1(s)$ sent $m_1 = s$. With such an alternative approach, a liar would consider the aggregate distribution of lie and (possibly) update it by conditioning on the information that $m_1 \neq s$. Clearly, the pure strategy equilibria shown in Proposition 1 would be unaffected by this alternative modeling to the extent that in such equilibria there is (at most) one lie s^h (and thus the extra conditioning has no bite for the lying party I). In appendix, I show that there cannot be pure strategy equilibria with multiple lies under this alternative modeling, thereby establishing the robustness of the analysis to such a variant.

4 Communicating with state-dependent objectives

I consider now situations in which party I's blisspoint action may depend on the state. Specifically, I let $u(a, s) = -(a - b(s))^2$ where $b(s)$ is assumed to be increasing with s. I wish to characterize the equilibria with forgetful liars as defined in Section 2 restricting attention to pure strategy equilibria.

The main observation is that multiple lies may arise in pure strategy equilibria when party I's objective is state-dependent. The key reason for this is that party I at $t = 2$, after a lie at $t = 1$, may end up choosing different messages as a function of the state despite having the same belief about what the first message was. This is so because the objective of party I is state-dependent and party I rightly anticipates which action is chosen by party U as a function of the messages. This, in turn, allows party I at $t = 1$ to safely engage in different lies as a function of the state, while still ensuring that he

will remain consistent throughout. Another observation concerns the structure of lies in equilibrium. I show that in all pure strategy equilibria, lies inducing larger actions a are associated with higher states, which eventually leads to a characterization of equilibria that borrow features both from cheap talk games (the interval/monotonicity aspect) and certification games (as seen in pure persuasion situations).

An example with multiple lies.

Assume that S consists of four equally likely states $s = 0$, s_1^*, s_2^* and 1. Let the bliss point function be $b(s) = s + \Delta$ for some Δ satisfying $\frac{1}{2} > \Delta > 0$.

I will look for conditions on s_1^*, s_2^* so that, in equilibrium, party I sends messages $m_1 = m_2 = s_1^*$ in states $s = 0$ and s_1^*, and party I sends messages $m_1 = m_2 = 1$ in states $s = s_2^*$ and 1.

In such a proposed equilibrium, party U must choose $a(s_1^*, s_1^*) = \frac{s_1^*}{2}$, $a(1,1) = \frac{s_2^*+1}{2}$ and $a(0,0) = 0$, $a(s_2^*, s_2^*) = s_2^*$ as well as $a(m_1, m_2) = 0$ for all other message profiles. With such strategies, two lies $m_1^* = s_1^*$ and $m_2^* = 1$ are made in equilibrium, and these two lies occur with the same probability. Thus, party $I_2(s)$ after a lie $m_1 \neq s$ at $t = 1$ believes at $t = 2$ that at $t = 1$ player $I_1(s)$ either sent $m_1 = s_1^*$ or 1, each with probability half.

To be an equilibrium, it should be that player $I_1(s_1^*)$ weakly prefers $a(s_1^*, s_1^*)$ to $a(1,1)$, as otherwise, player $I_1(s_1^*)$ would strictly prefer lying by sending $m_1 = 1$ anticipating that player $I_2(s_1^*)$ would also send $m_2 = 1$ (given that $I_2(s_1^*)$ would perceive that $m_1 = s_1^*$ or 1 are equally likely and inconsistent messages result in $a = 0$). Thus, $s_1^* + \Delta - a(s_1^*) \leq a(1) - s_1^* - \Delta$ or

$$\left(1 + s_1^* + s_2^*\right)/2 - 2\Delta \geq 2s_1^*. \tag{4}$$

More generally, it turns out that the incentives of $I_1(s)$ and $I_2(s)$ are aligned for all s. Thus, the remaining equilibrium conditions require that party I in state s_2^* weakly prefers $a(1)$ to $a(s_1^*)$ as otherwise, party I would strictly prefer the lie s_1^* to the lie 1 (both at $t = 1$ and 2). That is, $a(1) - s_2^* - \Delta \leq s_2^* + \Delta - a(s_1^*)$ or

$$2s_2^* \geq \left(1 + s_1^* + s_2^*\right)/2 - 2\Delta. \tag{5}$$

Moreover, it should be that party I in state $s = 0$ strictly prefers $a(s_1^*)$ to $a(0) = 0$ (what

he can get by telling the truth). That is, $a(s_1^*) < 2\Delta$ or

$$4\Delta > s_1^*. \tag{6}$$

Finally, it should be that party I in state $s = s_2^*$ strictly prefers $a(1)$ to $a(s_2^*) = s_2^*$ (what he can get by telling the truth). That is, $a(1) < s_2^* + 2\Delta$ or

$$4\Delta > 1 - s_2^*. \tag{7}$$

Whenever conditions (4)-(5)-(6)-(7) are satisfied (which is so whenever s_1^* is small enough and s_2^* is large enough, as soon as $\Delta < \frac{1}{2}$), the above two-lie communication strategy can be sustained as an equilibrium with forgetful liars. ♣

Characterization of equilibria employing pure strategies.

Pure strategy equilibria are characterized as follows where for any subsets A and B of S, I let $A < B$ whenever for all $s_A \in A$ and $s_B \in B$, we have that $s_A < s_B$.

No inconsistent messages are sent in equilibrium, as follows from an unravelling argument. Let m_k^* denote a consistent lie made by at least one type $s \neq m_k^*$ in equilibrium, and let L_k denote the set of types s such that party I with type s sends twice m_k^*, i.e. $m_1 = m_2 = m_k^*$. Let $L_k^- = L_k \backslash \{\max(s \in L_k)\}$ and $L = (L_k)_k$. Let $\widehat{a}_k(L) = E(s \in L_k)$ and $\left(\widehat{p}_k(L)\right)_k$ be such that $\widehat{p}_k(L)/\widehat{p}_{k'}(L) = p(L_k^-)/p(L_{k'}^-)$ (with $\sum_k \widehat{p}_k(L) = 1$). The following Proposition (proven in Appendix) summarizes the main properties of the pure strategy equilibria with forgetful liars.

Proposition 4 *There always exists an equilibrium with forgetful liars in pure strategies and any non-uniformly truthful such equilibrium satisfies the following properties. There is a disjoint family of lie sets $L = (L_k)_{k=1}^K$, with $L_1^- < \cdots < L_K^-$, $m_k^* = \max(s \in L_k)$ such that 1) Party I with type $s \in L_k^-$ lies twice by sending $m_1 = m_2 = m_k^*$; 2) Party I with type $s \in S \setminus \cup_k L_k^-$ tells twice the truth; 3) A liar's belief assigns probability $\widehat{p}_k(L)$ to $m_1 = m_k^*$; 4) Party U when hearing $m_1 = m_2 = m_k^*$ chooses $a = \widehat{a}_k(L)$; when hearing $m_1 = m_2 = s \in S \setminus \{m_1^*, ...m_K^*\}$ chooses $a = s$; and when hearing any other message profile chooses $a = 0$.*

In other words, lie sets L_k^- are ordered and the common lie in L_k^- is $m_k^* = \max(s \in L_k)$.

Party I in state s anticipates that if he lies at $t=1$ he will lie next by sending $m^*_{k(s)}$ where $k(s) = \arg\max_k \ v(k,s)$ and $v(k,s) = -\widehat{p}_k(L)(\widehat{a}_k(L) - b(s))^2 - (1-\widehat{p}_k(L))(a_{inc} - b(s))^2$ is party I 's time $t=2$ perceived expected utility of sending $m_2 = m^*_k$ after he lied at $t=1$ (the probability attached to $m_1 = m^*_k$ is $\widehat{p}_k(L)$ as follows from the consistency requirement (1)). To avoid being inconsistent, party I in state s will either send $m^*_{k(s)}$ both at $t=1$ and $t=2$ or he will be truthful (both at $t=1$ and $t=2$) depending on what he likes best.

Comment. When multiple lies m^*_k can be sustained in equilibrium, it is worth noting some similarity with the Perfect Bayes Nash equilibria that would arise in the one shot communication game in which all types except those corresponding to lies m^*_k could be certified (the similarity comes from the observation that types other than m^*_k either tell the truth (and get a payoff corresponding to the one they would get if they could fully disclose their type) or they consistently send message m^*_k).[30] Yet, a notable difference concerns the belief of a liar regarding which m^*_k he previously sent, which in turn induces incentive constraints typically more stringent than in the usual certification setup. Another difference already mentioned in the context of pure persuasion is that which type can be certified is endogenously determined by the equilibrium set of lies in the present context.

First-best with fine grid.

While multiple lies can arise in equilibrium when party I's objective may depend on the state, in the fine grid case (as defined in pure persuasion situations), it is not possible to sustain equilibria with multiple lies. Considering the general characterization shown in Proposition 4, in the fine grid case, all $\widehat{a}_k(L)$ must be approaching 0 as otherwise party I in too many states $s \in S$ smaller than $\widehat{a}_k(L)$ would be willing to make the lie m^*_k, making it in turn impossible to have that $\widehat{a}_k(L) = E(s \in L_k)$ (it is readily verified that there is only one state in L_k that lies above $\widehat{a}_k(L)$ and this is $s = m^*_k$). As a result, in the fine grid case, assuming that $b(s) \geq s + \Delta$ for some $\Delta > 0$, there can only be one lie in a pure strategy equilibrium, and the first-best for party U is being approached in the limit. This is similar to what was obtained in the pure persuasion case.

[30] Such a richer certification setup falls in the general framework defined in Green and Laffont (1986) or Okuno-Fujiwara and Postlewaite (1990).

5 Discussion

5.1 Back to criminal investigations

A key assumption driving the main insights is the memory asymmetry whether the informed party I lies or tells the truth at $t = 1$. With the criminal investigation application in mind, one may legitimately raise the concern that if a lying suspect pretends he is not guilty (i.e., by sending $m_1 = 1$ at $t = 1$) this may not be so hard to remember at $t = 2$, making the memory asymmetry assumption as considered in the main model not so clearly compelling in this case.

In most applications (criminal or otherwise), the full description of the state (or event) typically consists of many details, and not just a summary statistic, such as the level of guilt, that pins down parties' payoffs. When party I lies, he has to report details about the fabricated state, and it may be difficult later for party I to remember all these details when asked extra questions. In this case, it is not the overall level of guilt (as determined by the full description of the state) that is not remembered by the lying suspect, but the details reported in the lie. I would like to think of the main model as a simplified representation of such a richer specification. But, a question arises as to whether this view is legitimate and for what kind of communication protocols. Making progress on this may also be of interest to shed light on some experimental studies reported in the context of criminal investigation. In particular, Vrij et al (2008) has experimentally observed that when the communication protocol is too simple (for example always asking subjects to report the details in the same chronological order), subjects in the lab who are instructed to make lies tend to consistently report these details more correctly than when the communication protocol is less straightforward (for example asking subjects to report the details in reverse order). This may suggest that the type of communication protocol whether simple or non-straightforward may have implications on whether one may remain consistent when lying, which in turn may affect the incentive to lie in the first place.

While a full understanding of this would require further work, I would like now to propose a stylized modification of the main model that is suggestive of the directions such future research could take. Specifically, let me enrich the model as follows. Every state

now denoted θ consists of (s_A, s_B) where s_A and s_B assumed to be non-negative numbers correspond to the A and B attributes (or details) of the state θ, and $s = s_A + s_B$ summarizes the characteristics of the state (guilt level) parties I and U care about. As in Section 2, I assume that party U forms the best guess a about the expected value of s after the hearing of party I (she chooses action a and her objective is $-(a-s)^2$), and as in Section 3, party I who is informed of the state θ seeks to maximize a. There are finitely many states θ in Θ and the possible values of s are $s_1 = 0$,...$s_n = 1$ where s_k has probability $p(s_k)$. There is a small lying cost ε and the same genericity assumption as in the main model holds.

I will focus the analysis on a communication protocol that is clearly non-straightforward and I will then discuss how the analysis is modified when a simpler communication protocol is considered instead. Communication takes place at two times $t = 1, 2$. At $t = 1$, party I is asked to send a message m_1 describing the state either in normal order $\overrightarrow{m_1} = (\widehat{s}_A, \widehat{s}_B)$ or in reverse order $\overleftarrow{m_1} = (\widehat{s}_B, \widehat{s}_A)$ each with probability half. At $t = 2$, party I is asked to send a message about attribute X with $X = A$ (i.e. $m_2^A = \widetilde{s}_A$) or $X = B$ (i.e. $m_2 = \widetilde{s}_B$), each with probability half.[31] If the two messages are consistent (in the sense that $\widetilde{s}_X = \widehat{s}_X$) then party U is informed of $\widehat{s} = \widehat{s}_A + \widehat{s}_B$ and makes the best guess of s based on $\widehat{s}$, denoted $a(\widehat{s})$. To simplify the exposition of the arguments, I will be assuming that if the two messages are inconsistent, then party U is only informed of the inconsistency and chooses $a_{inc} = 0$ in such a case.[32]

Concerning party I's memory of m_1 at $t = 2$, I consider the following modification of the main model. As before, I distinguish the memory of party I at $t = 2$ about m_1 according to whether party I told the truth or lied at $t = 1$. If party I told the truth at $t = 1$, party I has perfect memory of m_1 at $t = 2$. If however party I at $t = 1$ lied, then party I at $t = 2$ has no memory of which $\widehat{s}_X$ for $X = A, B$ was reported. Party I's belief about $\widehat{s}_X$ is then the equilibrium aggregate distribution of first attribute (A or B)

[31] The possible request of describing the state in reverse order is in reference to some of the experimental settings considered in Vrij et al (2008). The randomization on the requested attribute is an extra level of complication with no clear link to Vrij et al' s work.

[32] Letting party U choose freely the action in case of consistency as well letting party U observe the exact choices of inconsistent messages would not affect the conclusions, but it would require extra analytical steps.

reported in m_1 when there was a lie at $t = 1$.[33] In all cases, party I remembers the state $\theta = (s_A, s_B)$.

The novelty compared to the main model is that after a lie at $t = 1$, party I is now only supposed to be confused (not remembering) the exact description of attribute X (A or B) in his message m_1. That is, unlike what was assumed in the main model, party I after a lie may remember the targeted level of guilt (as represented by $\widehat{s}_A + \widehat{s}_B$ in m_1).[34]

I will now sketch the main arguments why the pure strategy equilibria with forgetful liars in this modified setting take a form isomorphic to the ones shown in Proposition 1. I will then discuss why with other communication protocols -that should be thought of as simpler- or with alternative formalizations of forgetful liars (i.e. assuming a liar's belief about $\widehat{s}_X$ is conditional on $s = s_A + s_B$), other predictions may emerge.

Claim 1. The outcome of the one-lie equilibria of the main model as described by the (s^l, s^h)-communication strategy in Proposition 1 can be supported as equilibria with forgetful liars in the extended setting.

To see this, consider that in state $\theta = (s_A, s_B)$, party I sends $(\frac{s^h}{2}, \frac{s^h}{2})$ whether he is asked to report the state in normal order ($\overrightarrow{m_1}$) or in reverse order ($\overleftarrow{m_1}$) whenever $s = s_A + s_B \leq s^l$, and sends a truthful message m_1 ($\overrightarrow{m_1} = (s_A, s_B)$ or $\overleftarrow{m_1} = (s_B, s_A)$) otherwise. The trick of using such an attribute decomposition is that in this case, the aggregate distribution of $\widehat{s}_X$ conditional on a lie being made at $t = 1$ is a mass point on $\frac{s^h}{2}$. Thus, party I when lying at $t = 1$ will believe that he sent $\widehat{s}_X = \frac{s^h}{2}$ at $t = 1$ whether $X = A$ or B. To avoid being inconsistent, he will choose to send $m_2 = \frac{s^h}{2}$ at $t = 2$. As a result, those types who lie as just described will ensure they are consistent at $t = 2$ and thus induce the action $a(s^h)$ in equilibrium. If party I sends another lie at $t = 1$, i.e. $m_1 \neq (\frac{s^h}{2}, \frac{s^h}{2})$ (with m_1 being non-truthful), then at $t = 2$, party I will still report $m_2 = \frac{s^h}{2}$ no matter what X is (due to his belief about the false announced attribute),

[33] That is, aggregating for every state $\theta = (s_A, s_B)$ (with a weight proportional to the probability of θ), for every normal order request, $\widehat{s_A}$ whenever $\overrightarrow{m_1} = (\widehat{s}_A, \widehat{s}_B) \neq (s_A, s_B)$, and for every reverse order request, $\widehat{s}_B$ whenever $\overleftarrow{m_1} = (\widehat{s}_B, \widehat{s}_A) \neq (s_B, s_A)$.

[34] When asked about $\widehat{s}_X$, I am assuming that party I believes that $\widehat{s}_X$ is distributed according to the marginal aggregate distribution of $\widehat{s}_X$, as explained above. Party I could possibly refine this belief when remembering $\widehat{s}_A + \widehat{s}_B$, and realizing that the marginals of both $\widehat{s}_X$ and $\widehat{s}_{-X}$ (where $-X$ is attribute other than X) are the same. Such extra inferences are not trivial and they would become increasingly complex and weak if I were to consider a scenario with sufficiently many attributes. This leads me not to consider such inferences here.

and either for $X = A$ or B, party I will be inconsistent. This in turn deters party I from sending lies other than $(\frac{s^h}{2}, \frac{s^h}{2})$, and the remaining equilibrium conditions are easily verified.

Claim 2. There can be no pure strategy equilibria with forgetful liars admitting multiple lies.

To see this, observe that with multiple lies, the support of the equilibrium distribution of $\widehat{s}_X$ conditional on a lie being made would have to contain at least two different values (the trick used for claim 1 cannot work for all lies if there are different levels of targeted guilt). Given that at $t = 2$, the belief of a liar about $\widehat{s}_X$ would be the same whether $X = A$ or B, party I would send the same message m_2 whether asked to report attribute $X = A$ or B (this is making use of the genericity assumption). As a result, party I for at least one lie and one realization of X would be inconsistent. Party I would prefer avoiding this by being truthful throughout, thereby explaining why it is not possible to support equilibria with multiple lies in this modified setting.

Comments.

1. As in the main model, in the fine grid case, all equilibria employing pure strategies result in the almost perfect elicitation of the state.

2. If one assumes that party I knows the distribution of $\widehat{s}_X$ conditional on θ when a lie is being made at $t = 1$ (as in the approach similar to Piccione and Rubinstein discussed above), then many more lies can be supported in pure strategy equilibria (when only sending $m_1 = (\frac{\widehat{s}}{2}, \frac{\widehat{s}}{2})$ in state θ, party I can ensure not being inconsistent in such a variant). As in Subsection 3.4.1, in this case, one cannot expect the full elicitation of the state, even in the fine grid case.

3. If one were to modify the communication protocol and assume instead that party I at $t = 2$ is always asked to report the A attribute (instead of randomizing between attributes A and B), one could again support many more lies in pure strategy equilibria. This, for example, can be seen by assuming that whenever party I lies, he chooses always $\widehat{s}_A = 0$ while adjusting $\widehat{s}_B = \widehat{s}$ to the targeted level of guilt $\widehat{s}$. In this case, 0 is the dominant mode in the aggregate distribution of lies, thereby ensuring that at $t = 2$ after a lie at $t = 1$, party I would always report that the A attribute is 0. By choosing $\widehat{s}_A = 0$ at $t = 1$, party I could safely avoid being inconsistent and strategize as if he had perfect

memory. Such an insight together with the analysis of the more complex communication protocol in which the requested attribute X at $t = 2$ is randomized gives some theoretical support to the experimental finding of Vrij et al. (2008).[35]

5.2 When some liars are caught being inconsistent

Except for the equilibria in mixed strategy discussed in Subsection 3.3, the equilibria with forgetful liars as characterized above have the feature that there are no inconsistent messages sent in equilibrium. Being caught sending inconsistent messages is detrimental, and in equilibrium this disciplines the informed party I into never sending inconsistent messages. While it would be desirable to enrich the model so that inconsistent messages can happen (in some equilibria employing pure strategies), I note that this is unlikely to arise when being inconsistent has detrimental consequences (which sounds like a natural feature) and party I is viewed as having rational expectations at $t = 1$ and at later time periods, as long as there is no lie. Indeed, party I can avoid being inconsistent simply by telling the truth throughout. When party I has rational expectations as long as not lie has been made, this option ensures that no inconsistency can happen in any pure strategy equilibrium when the outcome in case of inconsistency is no better than the outcome in case of truth-telling. This simple argument reveals that to obtain some inconsistency in a pure strategy equilibrium (while inconsistency continues to have detrimental consequences), one would need to relax the rational expectation assumption beyond the events in which some lie has already been made. Thus, cognitive errors beyond those implied by the memory imperfection of liars should be considered. Along such lines, one could for example consider cognitive environments in which sometimes party I would have incorrect beliefs at $t = 1$ about the message that would be sent at $t = 2$ after a lie at $t = 1$ (maybe using the aggregate distribution over all time $t = 2$ messages to form their belief). With such an extra feature of bounded rationality, party I could find it subjectively good to be lying at $t = 1$ (based on his erroneous expectation), and party I could unexpectedly be caught being inconsistent later on. Clearly, more work is needed to

[35] In a very different context, Glazer and Rubinstein (2014) also suggest in a theoretical framework how complex questionnaires may help elicit the truth when the informed party faces constraints. Yet, the constraints considered in Glazer and Rubinstein cannot directly be related to memory asymmetries as considered in this paper.

formalize this completely as well as to find compelling formulations of such extra features of bounded rationality.

5.3 Some further theoretical considerations

I will discuss two items here. The first concerns whether one can always view the equilibria with forgetful liars as defined in Section 2 as selections of equilibria with imperfect recall as discussed in Subsection 3.4.1 in which party I would be assumed to know how his lying strategy varies with the state. The second investigates the possibility that party U would be able to commit to some pre-specified course of action (as a function of the outcome of the communication).

Do equilibria with forgetful liars remain equilibria when liars remember their strategy?

Restricting attention to pure strategy equilibria in the main model, it can be checked that the one-lie equilibria with forgetful liars can also be viewed as equilibria with imperfect recall in which liars would know how their lying strategy depends on the state and party I in a state s where he is supposed to tell the truth would believe at $t = 2$ if lying at $t = 1$ that he lied according to the unique lie made in equilibrium (in other states $s' \neq s$). That is, the trembling required to support the equilibria with forgetful liars as equilibria with imperfect recall would have to be degenerate (mass point on one message) and equilibrium-specific (the unique lie made by others in equilibrium). If one were to exogenously impose some trembling that would not be degenerate and/or would be fixed independently of the equilibrium, there is no reason to expect the equilibria with forgetful liars to be equilibria in the imperfect recall sense.

As a further elaboration on the difference between the two approaches, consider in the state-dependent objective scenario, a setting in which a pure strategy equilibrium with forgetful liars would have multiple lies, and to fix ideas consider the example provided in Section 4. In this setting, the belief at $t = 2$ of party I in state $s = s_2^*$ is that he either sent $m_1 = s_1^*$ or 1, each with probability half at $t = 1$. When party I knows how his time $t = 1$ strategy depends on s, party I would at $t = 2$ know he sent $m_1 = 1$ at $t = 1$, resulting in a different belief of party I. In the context of the game as considered in the main model, the optimal behavior at $t = 2$ of party I in state $s = s_2^*$ would still be to send

$m_2 = 1$ with such a correct belief, thereby ensuring that the strategy profile considered in the equilibrium with forgetful liars is also an equilibrium with imperfect recall in which the liar would know how his strategy varies with s.

But, the difference in liars' beliefs can have bigger consequences in more complex communication protocols. For the sake of illustration, consider a variant of the main communication game in which at $t = 2$, sometimes with some positive probability, party I is given the opportunity to confess that he lied, resulting then in an action not too far from s_2^*. If the opportunity to confess is small enough, not much of the analysis is affected except that now at $t = 2$ party I in state s_2^* will choose to confess whenever possible because given his belief of what message he sent at $t = 1$ he attaches a (subjective) probability 0.5 that he may be declared inconsistent (resulting in $a = 0$) if he reports $m_2 = 1$ instead of confessing. By contrast, if party I knows his strategy, he would not confess, as he would rightly believe he sent $m_1 = 1$ at $t = 1$ in this event, thereby making the confess option an unattractive one. In this case, the equilibrium with forgetful liar is not an equilibrium with imperfect recall no matter how the trembles are defined.

Mechanism design and commitment

Suppose in the context of the communication game as described in Section 2 that party U could commit in advance to choosing some action $a(m_1, m_2)$ when messages m_1 and m_2 are sent at $t = 1, 2$ (while party I 's memory problems would be modeled in the same way as in Section 2).

Clearly, any specific equilibrium with forgetful liars as described in Sections 3 and 4 can be obtained as an equilibrium in the commitment world by assuming that $a(m_1, m_2)$ for the various (m_1, m_2) are set as in the corresponding equilibrium. A more interesting observation is that fixing $a(m_1, m_2)$ as in one such equilibrium may now generate more equilibria of the communication game in the commitment world. As it turns out, no matter how $a(m_1, m_2)$ for the various possible messages m_1, m_2 are set, it may be that some equilibria in the commitment world remain bounded away from the first-best, even in the limit as the grid gets finer and finer. Such a conclusion would not arise in classic certification games.[36] It is suggestive that there may be a potential benefit of the absence

[36] Consider Milgrom's certification game. There, all types can be certified, and if party U commits to choosing the worst possible action if the state is not disclosed, only the first-best arises, exactly as in the

of commitment in environments with forgetful liars.[37]

To see this, consider the pure persuasion case, and assume that $a(1,1)$ is set close to 1 while $a(s,s)$ is set below 1 and $a(m_1,m_2)=0$ for all other message profiles (m_1,m_2) (as should be the case if one wishes to approach the first-best in the fine grid case). One equilibrium with forgetful liars in the induced game with such a committed party U is that whatever the state s, party I sends twice $m_1=m_2=1$ resulting in action $a(1,1)=1$ for all states (which is clearly far away from the first-best).

Indeed, with this communication strategy in place, all lies are concentrated on 1. Hence, when party I in state $s\neq 1$ lies and sends $m_1=1$ at $t=1$, he can safely anticipate he will choose $m_2=1$ at $t=2$ (so as to avoid being inconsistent). This strategy results in action $a(1,1)=1$, and this strategy is optimal given that $a(1,1)=1$ is larger than 0 (the action that would result if party I were to send another lie at $t=1$) and $a(s,s)$ if party I in state $s\neq 1$ were telling the truth throughout. The difference with the analysis of the game of Section 2 is that now party U does not react to the chosen equilibrium (in the proposed strategy of party I, party U would have chosen $a(1,1)=E(s)$ in the context of the main model while now she is committed to choosing $a(1,1)=1$) and this lack of reaction of party U in turn causes the emergence of many more equilibria including ones that are suboptimal from party U's perspective.

6 Conclusion

This paper has offered an analysis of how multi-round communication protocols may help elicit considerable information with forgetful liars. While I have included a discussion of several alternative modeling of forgetful liars as well as their implications in terms of communication strategies, many additional extensions deserve extra work. These include the modeling of partial memories of lies, the considerations of richer contexts in which

game without commitment. A similar comment applies to Dye's setting.

[37] There have been some recent papers (see in particular Ben Porath et al. (2019), Hart et al. (2016) or Sher (2011)) starting with Glazer and Rubinstein (2004) that establish in various persuasion environments that commitment of the uninformed party may be unnecessary. The insight developed by these papers is that the best outcome achievable through a mechanism with full commitment can be attained as one equilibrium of the game without commitment. It thus follows a weak implementation perspective in contrast with the full implementation perspective suggested here.

the informed party may have different knowledge of the state at different times, a different memory treatment for ordinary lies and lies by omission as well as richer cognitive environments allowing for the emergence of inconsistency in pure strategy equilibria.

Appendix

Proof of Lemma 1

Call β the aggregate distribution of lies at $t = 1$. In a pure strategy equilibrium, it takes the form that for a family $\left\{m^k\right\}_k$ of messages, m^k is assigned probability p^k where p^k is proportional to $p(T^k)$ and T^k is the subset of S such that $\sigma_1(s) = m^k$ and $s \neq m^k$.

At $t = 2$, after a lie at $t = 1$, player $I_2(s)$ will assess that sending m_2 would give an expected continuation payoff equal to

$$\left(\sum_k p(T^k)a(m^k, m_2)\right) / \left(\sum_k p(T^k)\right) - \varepsilon 1_{\{m_2 \neq s\}}.$$

Given that in a pure strategy equilibrium if (m^k, m_2) is on the path, we must have that $a(m^k, m_2) = E(s \in T^k(m_2))$ where $T^k(m_2)$ is the set of s such that $\sigma_1(s) = m^k$ and $\sigma_2(s) = m_2$ and that any two different m_2, m_2' would induce different $T^k(m_2)$, $T^k(m_2)$, the requirement on ε being small enough implies that whatever s, there is a unique best-response $m^* \in \left\{m^k\right\}_k$.[38] ♣

Proof of Lemma 2. Suppose by contradiction that $m^* \neq \max\{s \in L(m^*)\}$ and let $s^* = \max\{s \in L(m^*)\}$. Player $I_1(s^*)$ when sending $\sigma_1(s^*) = m^*$ should expect to get $E(s \in L(m^*)) - 2\varepsilon$ (anticipating that player $I_2(s^*)$ will tell m^* as follows from Lemma 1). But, if player $I_1(s^*)$ deviates and tells the truth, player $I_2(s^*)$ would know this and could decide to tell the truth at $t = 2$. Thus, when player $I_1(s^*)$ is truthful at $t = 1$, both players $I_1(s^*)$ and $I_2(s^*)$ would get a payoff no smaller than $a(s^*, s^*)$. Given that by Lemma 1, anyone lying at $t = 1$ must be sending m^* at $t = 2$, it follows that the message profile $(m_1 = s^*, m_2 = s^*)$ would be off-the-path so that one should have $a(s^*, s^*) = s^*$. Given

[38] The best-response cannot be outside $\left\{m^k\right\}_k$ unless all $a(m^k, m^{k'}) = 0$ which would then imply that only $s = 0$ is lying at $t = 1$ and this would lead to a contradiction as then player $I_1(s = 0)$ would be strictly better off telling the truth at $t = 1$.

that $E(s \in L(m^*)) \leq s^*$, we conclude that the deviation would be profitable, thereby leading to a contradiction. ♣

Proof of Lemma 3. By Lemma 2, if $\sigma_1(s) = s$ and $\sigma_2(s) \neq s$, then $s \neq m^*$. Moreover if $\sigma_2(s) \neq m^*$, then $(m_1 = s, m_2 = \sigma_2(s))$ would perfectly reveal the state s to party U as no other type could be using such a communication strategy by Lemma 1. If $I_2(s)$ deviates and tells the truth, he would induce action s (as (s, s) would be off-the-path as follows from Lemma 1) and save the lying cost ε, thereby making the deviation to truth-telling profitable. ♣

Proof of Lemma 4. Suppose by contradiction that $S_{inc}(m, m^*) \neq \varnothing$ and let $s^*_{inc} = \max S_{inc}(m, m^*)$. We know that $s^*_{inc} \neq m^*$ since $m^* \in L(m^*)$ and $L(m^*) \cap S_{inc}(m, m^*) = \varnothing$. Player $I_2(s^*_{inc})$ anticipates to get at most $E(s \in S_{inc}(m, m^*)) - \varepsilon$ by following his assumed equilibrium strategy. Assuming $m \neq s^*_{inc}$, I show that player $I_1(s^*_{inc})$ can strictly gain by telling the truth. In such a case, players $I_1(s^*_{inc})$ and $I_2(s^*_{inc})$ would secure a payoff at least as large as what results when $m_1 = m_2 = s^*_{inc}$ are sent. But such a message profile would be off-path by the observation that $s^*_{inc} \neq m^*$ and thus s^*_{inc} cannot be a lie made at $t = 2$, thereby implying that $m_1 = m_2 = s^*_{inc}$ results in action s^*_{inc}. Thus player $I_1(s^*_{inc})$ would be strictly better off telling the truth, thereby leading to a contradiction. Assuming instead that $m = s^*_{inc}$ would lead player $I_2(s^*_{inc})$ to strictly prefer telling the truth rather than m^*, leading again to a contradiction. ♣

Proof of Proposition 2

Let a^*_n denote the equilibrium action after a lie in σ^n. Suppose by contradiction that for some $\widehat{a}$ and all $n > \overline{n}$, $a^*_n > \widehat{a}$. There must be at least $n\widehat{a}/2$ states s_k smaller than a^*_n in S_n. Moreover, the fine grid assumption implies that $E(s \in S_n,\, s < a^*_n) < a^*_n - \underline{\alpha}\widehat{a}/2(\underline{\alpha}+\overline{\alpha})$ for n large enough. Moreover, for n large, we would have $\Pr(s < a^*_n) > n\underline{\alpha}\widehat{a}/2(\underline{\alpha} + \overline{\alpha})$, thereby implying that $E(s \text{ such that } s < a^*_n \text{ or } s = 1) < a^*_n - \underline{\alpha}\widehat{a}/3(\underline{\alpha} + \overline{\alpha})$ (making it impossible to meet the requirement $a^E(s^l, s^h) - 2\varepsilon \geq s^l$ with s^h close to a^*_n. This leads to inconsistent conditions, thereby showing the desired result. ♣

Pure strategy equilibria when liars remember that they lied

Suppose there are several lies m^*_k made in equilibrium.

For each k, define $L_k = \{s \text{ such that } \sigma_1(s) = m_k^*\}$.

One should have that $\overline{s}_k = \max L_k$ is such that for some k', $\sigma_1(\overline{s}_k) = \sigma_2(\overline{s}_k) = m_{k'}^*$ as otherwise players $I_t(\overline{s}_k)$ would strictly prefer sending $m_t = \overline{s}_k$ (unravelling argument). In fact, the lying costs would then imply that $\sigma_1(\overline{s}_k) = \sigma_2(\overline{s}_k) = \overline{s}_k$ (as otherwise one of the $\overline{s}_k$ supposed to be lying would strictly prefer telling the truth). Inconsistent messages can also be ruled out by an unravelling argument.

This implies that for every k, L_k contains more than one state and that every $s \in L_k \setminus \{\overline{s}_k\}$ must be different from $m_{k'}^*$ for all k'. The optimality of the strategy of $I_2(s)$ for $s \in L_k \setminus \{\overline{s}_k\}$ would then imply that

$$p(L_k)a(m_k^*, m_k^*) = \max_{k'} p(L_k)a(m_k^*, m_k^*)$$

which in turn by the genericity assumption implies that there can be only one lie and that the analysis of Proposition 1 applies. ♣

Proof of Proposition 4

Let m_k^* denote a consistent lie made by at least one type $s \neq m_k^*$, i.e. party I with type s sends twice the message m_k^*, and assume there are K different such lies in equilibrium. Define then L_k as the set of types s such that party I with type s sends twice m_k^*, i.e. $m_1 = m_2 = m_k^*$ (this includes those types who lie and say consistently m_k^* and possibly type $s = m_k^*$ if this type tells the truth), and let $L = (L_k)_k$. Clearly, in such an equilibrium, after the message m_k^* has been sent twice, party U would choose $a_k = E(s \in L_k)$. I let $\overline{s}_k$ denote $\max L_k$ and observe that $\overline{s}_k$ should be one of the consistent lies m_r^* for $r = 1..., K$:

Lemma 5 *For all k, $\overline{s}_k = \max L_k$ should be a consistent lie.*

Proof. Suppose this is not the case. Then party I with type $\overline{s}_k$ would induce action $a = \overline{s}_k$ by telling twice the truth. This would be strictly better for him than what he obtains by sending twice m_k^*, which gives action $a_k = E(s \in L_k) \leq \overline{s}_k = \max L_k$ (and inflicts an extra 2ε penalty for not telling the truth - this is needed to take care of the case in which L_k would consist of $\overline{s}_k$ only). ♣

A simple implication of lemma 5 is:

Corollary 1 *There is a bijection between* $\{L_1, ...L_K\}$ *and* $\{\overline{s}_1, ...\overline{s}_K\}$.

Another observation similar to that obtained in pure persuasion games is:

Lemma 6 *There can be no (voluntary) inconsistent messages sent by any type* $s \neq 0$ *in equilibrium.*

Proof. Let $S_{inc}(m_1, m_2) = \{s \text{ such that } \sigma_1(s) = m_1 \text{ and } \sigma_2(s) = m_2\}$ with $m_1 \neq m_2$ and assume by contradiction that $S_{inc}(m_1, m_2) \neq \varnothing$. By Corollary 1, on can infer that $m_k^* \notin S_{inc}(m_1, m_2)$. Let $s_{inc}^*(m_1, m_2) = \max S_{inc}(m_1, m_2)$. It is readily verified that $I_1(s_{inc}^*(m_1, m_2))$ and $I_2(s_{inc}^*(m_1, m_2))$ are strictly better off telling the truth, thereby leading to a contradiction. ♣

Let μ_k denote the overall probability (aggregating over all s) with which m_k^* is sent at $t = 1$ conditional on a lie being sent then (i.e., conditional on $m_1 \neq s$). Without loss of generality reorder the k so that $\mu_k a_k$ increases with k. The single crossing property of $u(a, s)$ implies that:

Lemma 7 *For any* $k_1 < k_2$*, if in equilibrium* $I(s)$ *lies by sending twice* $m_{k_1}^*$ *and* $I(s')$ *lies by sending twice* $m_{k_2}^*$*, it must be that* $s < s'$*. Moreover, for every* k*, it must be that the consistent lie* m_k^* *in* L_k *coincides with* $\max L_k$*, i.e.* $\overline{s}_k = m_k^*$.

Proof. For the first part, note that after a lie, player $I_2(s)$ would send $m_2 = m_{k(s)}^*$ where

$$\begin{aligned} k(s) &= \arg\max_k v(k, s) \text{ and} \\ v(k, s) &= -\mu_k(a_k - b(s))^2 - (1 - \mu_k)(a_{inc} - b(s))^2. \end{aligned}$$

Given that $a_{inc} = 0$, and $\mu_1 a_1 < \mu_2 a_2 ... < \mu_K a_K$ (they cannot be equal by the genericity assumption), it is readily verified that for any $s_1 < s_2$, and $k_1 < k_2$, if $v(k_2, s_1) > v(k_1, s_1)$ then $v(k_2, s_2) > v(k_1, s_2)$.[39]

Thus if party I with type s_2 finds lie $m_{k_2}^*$ optimal, he must find it better than $m_{k_1}^*$ and thus by the property just noted, party I with any type $s > s_2$ must also find $m_{k_2}^*$ better than $m_{k_1}^*$, making it impossible that he finds $m_{k_1}^*$ optimal.

[39] This makes use of $(v(k_2, s_2) - v(k_1, s_2)) - (v(k_2, s_1) - v(k_1, s_1)) = 2(\mu_{k_2} a_{k_2} - \mu_{k_1} a_{k_1})(b(s_2) - b(s_1))$ noting that $b(s_2) > b(s_1)$.

To show the second part ($\overline{s}_k = m_k^*$), I make use of Corollary 1 to establish that if it were not the case there would exist an increasing sequence $k_1 < k_2 ... < k_J$ such that type $\overline{s}_{k_j}$ would lie by sending $\overline{s}_{k_{j+1}}$ for $j < J$ and $\overline{s}_{k_J}$ would lie by sending $\overline{s}_{k_1}$, which would violate the property just established. ♣

To complete the description of equilibria, let $L_k^- = L_k \setminus \{m_k^*\}$ where $m_k^* = \overline{s}_k = \max L_k$; $p(L_k^-)$ denote the probability that $s \in L_k^-$; $\mu_k(L) = p(L_k^-)/\left(\sum_r p(L_r^-)\right)$ the probability that the lie m_k^* is made at $t = 1$ in the aggregate distribution of lies at $t = 1$; $k(s) = \arg\max_k v(k, s)$ where $v(k, s) = -\mu_k(L)(a_k - b(s))^2 - (1 - \mu_k(L))(b(s))^2$ and $a_k(L) = E(s \in L_k)$. Realizing that party I with a type s that lies outside $\{m_1^*, ...m_K^*\}$ will either tell the truth or lie by sending $m_{k(s)}^*$ depending on what he likes best, and that by Lemma 7 party I with type $\overline{s}_k = m_k^*$ should prefer telling the truth to lying by sending $m_{k(\overline{s}_k)}^*$, the conditions shown in Proposition 4 follow.

Finally, to show that there exists an equilibrium in pure strategies with some lying activity, think of having a unique lie set, $K = 1$, and let $L_1 = \{s_1, s_2\}$ with the lie being $m_1^* = s_2$, and consider the strategies as specified in the proposition. It is readily verified that all the required conditions are satisfied.♣

References

[1] Aumann Robert and Sergiu Hart (2003): 'Long cheap talk,' *Econometrica* 71, 1619–1660.

[2] Balbuzanov Ivan (2017): "Lies and consequences: The effect of lie detection on communication outcomes," mimeo.

[3] Ben-Porath Elhanan, Eddie Dekel, and Barton Lipman (2019): 'Mechanisms with evidence: Commitment and robustness,' *Econometrica* 87, 529-566.

[4] Crawford Vincent, and Joel Sobel (1982): 'Strategic information transmission,' *Econometrica* 50, 1431–1451.

[5] Chen Ying (2011): 'Perturbed communication games with honest senders and naive receivers,' *Journal of Economic Theory* 146, 401–424

[6] Crawford Vincent (2003): 'Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions,' *American Economic Review* 93, 133-149.

[7] Deneckere Robert and Sergei Severinov (2017): 'Screening, signalling and costly misrepresentation,' mimeo.

[8] Dye Ra (1985): 'Strategic Accounting Choice and the Effects of Alternative Financial Reporting Requirements' *Journal of Accounting Research* 23(2): 544–74.

[9] Dziuda Wioletta. and Christian Salas (2018): "Communication with detectable deceit," mimeo.

[10] Ettinger David and Philippe Jehiel (2010): 'A theory of deception,' *American Economic Journal: Microeconomics* 2, 1-20.

[11] Forges Françoise. (1990): 'Equilibria with communication in a job market example,' *Quarterly Journal of Economics* 105, 375–398.

[12] Glazer Jacob, and Ariel Rubinstein (2004): "On Optimal Rules of Persuasion," *Econometrica* 72(6): 1715–36.

[13] Glazer Jacob, and Ariel Rubinstein (2006): "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach," *Theoretical Economics* 1 395–410.

[14] Glazer Jacob, and Ariel Rubinstein (2014): 'Complex questionnaires,' *Econometrica* 82, 1529–1541.

[15] Gneezy Uri (2005): 'Deception: The role of consequences,' *American Economic Review* 95, 384-394.

[16] Green Jerry and Nancy Stokey (2007): 'A two-person game of information transmission,' *Journal of Economic Theory* 135, 90–104

[17] Green Jerry and Jean-Jacques Laffont (1986): 'Partially Verifiable Information and Mechanism Design,' *Review of Economic Studies* 53 (3): 447–56.

[18] Grossman Sanford (1981) "The Informational Role of Warranties and Private Disclosure about Product Quality." *Journal of Law and Economics* 24 (3): 461–83.

[19] Grossman Sanford, and Oliver Hart. (1980): "Disclosure Laws and Takeover Bids." *Journal of Finance* 35 (2): 323–34.

[20] Hart Sergiu, Ilan Kremer and Motty Perry (2017): 'Evidence Games: Truth and Commitment,' *American Economic Review* 107, 690-713.

[21] Hörner Johannes, Xiaosheng Mu, and Nicolas Vieille (2017): "Keeping your story straight: Truthtelling and liespotting," mimeo.

[22] Jehiel Philippe (2005): 'Analogy-based expectation equilibrium,' *Journal of Economic Theory* 123, 81-104.

[23] Jehiel Philippe (2019): 'Communication with forgetful liars,' Working paper *Paris School of Economics.*

[24] Jehiel Philippe and Frédéric Koessler (2008): 'Revisiting Bayesian games with analogy-based expectations,' *Games and Economic Behavior* 62, 533-557.

[25] Kartik Navin (2009): 'Strategic communication with lying costs,' *Review of Economic Studies* 76, (4), 1359-1395

[26] Kartik Navin, Marco Ottaviani, and Francesco Squintani (2007): 'Credulity, lies, and costly talk,' *Journal of Economic Theory* 134, 93 – 116.

[27] Krishna Vijay, and John Morgan (2004): 'The Art of Conversation: Eliciting Information from Experts through Multi-Stage Communication," *Journal of Economic Theory* 117, 147-179.

[28] Milgrom Paul (1981): 'Good News and Bad News: Representation Theorems and Applications,' *Bell Journal of Economics* 12 (2): 380–91.

[29] Okuno-Fuijiwara Masahiro and Andrew Postlewaite (1990): 'Strategic Information Revelation,' *Review of Economic Studies* 57(1), 25-47.

[30] Piccione Michele and Ariel Rubinstein (1997): 'On the Interpretation of Decision Problems with Imperfect Recall,' *Games and Economic Behavior* 20, 3-24.

[31] Sher Itai (2011): 'Credibility and determinism in a game of persuasion,' *Games and Economic Behavior* 71, 409-419.

[32] Sobel Joel (2018): "Lying and deception in games," forthcoming *Journal of Political Economy.*

[33] Vrij Aldert, Samantha Fisher, Ronald Leal, Sharon Milne, Rebecca and Bull Ray (2008): 'Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order,' *Law and Human Behavior* 32, 253-265.

[34] Vrij Aldert, Pär Anders Granhag, Samantha Mann, and Sharon Leal (2011): 'Outsmarting the liars: Toward a cognitive lie detection approach,' *Current Directions on Psychological Science* 20(1), 28-32.