Communication with Forgetful Liars

Philippe Jehiel

WORKING PAPER N° 2019 – 37

Communication with Forgetful Liars

Philippe Jehiel

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

Communication with Forgetful Liars*

Philippe Jehiel†

8th July 2019

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, when multiple lies can arise in equilibrium, and when inconsistencies trigger harmful consequences.

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

*I wish to thank 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, and ESSET 2018 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. Very often, 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 (see the discussion and literature review in Vrij et al. (2011)).[1] The objective of this paper is to develop a game theoretic framework that formalizes 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 beyond the criminal investigation application to the extent that the memory asymmetry between liars and truth-tellers seems widespread. 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 Sobel (1982)). But, how is this insight affected in the

[1] It should be mentioned that the idea that lies may be hard to remember appears also in the popular culture. For example, it is subtly expressed in a quote attributed to Mark Twain as "*When you tell the truth you do not have to remember anything,*" which implicitly but clearly suggests a memory asymmetry whether you tell the truth or you lie.

presence of forgetful liars?

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 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 $s = 1$ will be interpreted as complete innocence and $s = 0$ as full guilt. After hearing the outcome of the communication phase, 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, but if he lies by saying $m_1 \neq s$, 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. When two identical messages $m_1 = m_2 = m$ are being sent by party I at $t = 1$ and 2, party U observes the message m, but when m_1, m_2 with $m_1 \neq m_2$ are being sent, I assume that party U is only informed of the inconsistencies (i.e., that $m_1 \neq m_2$). Party U is assumed to make the optimal choice of action given what she is told.[4]

[2]Such an assumption is -I believe- natural in contexts in which the record of past communication interactions would highlight the type of lies that were used rather than the joint description of the lie and the event s (the details of which are typically disclosed with a different time scale). While this will be my main modelling approach, I will also discuss the implications of an alternative approach in which party I would be viewed as knowing how his communication strategy depends on s.

[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).

[4]The assumption that only inconsistency is observed by party U when $m_1 \neq m_2$ allows me to simplify

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 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, with a tiny probability, party I always communicates the truth and party I incurs a tiny extra cost when lying (so that party I would consider lying only if it is strictly beneficial).[5] 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 properties. Either party I always tells the truth or there is exactly one lie made in equilibrium. In the latter case, calling s^h the unique lie, 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 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 equilibria with forgetful liars whether in pure or in mixed strategies lead approximately to the first-best in which party U perfectly infers the state whatever s and chooses the action $a = s$ accordingly.

Thus, in pure persuasion situations, my analysis reveals that with forgetful liars, simple multi-round communication protocols ensure that party U obtains much more information from party I as compared with one-shot communication protocols in which party I would not reveal any information (as results from the standard analysis without memory imperfections). Moreover, when there is some significant lying activity (i.e. moving away from the fine grid case), there is only one lie occurring in a pure strategy equilibrium, and

some of the analysis, but it is not needed for the derivation of the main insights reported below. I also believe it fits with a number of applications in which the decision maker does not directly participate in the hearings but is presented with a summary of those.

[5] I also assume in the pure persuasion case that faced with the same expectations, party I behaves in the same way irrespective of the state, which can be rationalized by considering (small) state-independent idiosyncratic preference perturbations. The role of such perturbations, which I believe are plausible in most applications, will be discussed later on.

this unique lie is made only for low levels of s. I also note that in all equilibria whether in pure or in mixed strategy, the expected utility obtained by party I is no smaller than what party I would obtain by being inconsistent, thereby endogenizing the costly nature of being inconsistent.

I next explore how the analysis is affected when the objective of the informed party I may depend on the state s as in Crawford-Sobel's cheap talk games. The main observation is that then multiple lies can arise in equilibria employing pure strategies. The reason is as follows. After a lie at $t = 1$, party I at $t = 2$ may now opt for different m_2 depending on the state s because party I rightly understands how party U's action varies with m (when $m = m_1 = m_2$) 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 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 attributed 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 the experimental finding reported in Vrij et al. (2008) who advocate in favor of increasing cognitive load so as to facilitate lie detection.

Related Literature

The above findings can be related to different strands of literature. First, there is a large literature on cheap talk as 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. 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 this is however unrelated to the memory imperfections considered in this paper, and for example the insight obtained in this paper that the first-best is approached in the equilibria with forgetful liars in the fine grid case has no counterpart in that literature.[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 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))

[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, in particular, Forges (1990), Aumann and Hart (2003) or Krishna and Morgan (2004)).

[7] Interestingly, Mark Twain's quote as reported in footnote 3 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.

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, even in the fine grid case.

Fourth, it may be interesting to compare the results obtained here with those obtained when explicit lying costs (possibly determined by the distance between the state and the lie) are added to the standard cheap talk game (as in Kartik (2009)). In the case of lying costs, every type has an incentive to inflate his type and there is some pooling at the highest messages, which sharply contrasts with the shape of the equilibria with forgetful liars in pure persuasion situations in which pooling occurs for low types.[8]

Finally, Dziuda and Salas (2018) 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]

[8] In a mechanism design setting, Deneckere and Severinov (2017) assume that each time the informed party misreports his type, he incurs an extra cost. They make the observation that in such a setting, 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 assumptions 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 setting 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

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.

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 $s_n = 1$. The ex ante probability that the realized state s_k arises is $p(s_k)$, which is commonly known, and $S = \{s_k\}_{k=1}^n$ denotes the state space. 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

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.

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 $M \setminus S \neq \varnothing$. Second, in order to have 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. When the two messages are the same $m_1 = m_2 = m$, party U is informed of m. When they are inconsistent in the sense that $m_1 \neq m_2$, party U is only informed that $m_1 \neq m_2$. In all cases, party U chooses her action a based on what she is told about the communication phase. That is, $a(m)$ if $m_1 = m_2 = m$, and a_{inc} if $m_1 \neq m_2$.[10]

Having 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. says $m_1 = s$, he remembers that $m_1 = s$ at $t = 2$, but when he lies (identified here with not telling the whole truth) and says $m_1 \neq s$, 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 further discussion on this below).

A key modeling issue is about 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 solution concept, referred to as equilibrium with forgetful liars.

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 preferences given by (2). To model the belief of a forgetful

[10] I have formulated the communication phase as one in which party U would not be present and would only be informed of some aspects of it when $m_1 \neq m_2$. An alternative interpretation is that party U would be constrained to choose the same action when $m_1 \neq m_2$, which can be motivated on the ground that outside parties who judge party U would only be informed that there were inconsistencies in such a case (while being informed of the sent messages when consistent). I also briefly argue later on that the main insights carry over if party U observes (m_1, m_2) even if $m_1 \neq m_2$.

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. says $m_1 \neq s$), player $I_2(s)$ at time $t = 2$ believes that player $I_1(s)$ sent m with probability (1). If no lie is ever made 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., says $m_1 = s$), player $I_2(s)$ knows that $m_1 = s$. That is, truth-tellers remember they told the truth.

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 (some may correspond to off-path observations in which case they are free), 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).

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

Perturbations.

First, I assume that in case of indifference between lying and truth-telling, party I opts for truth-telling. Also, to simplify some of the arguments, I assume that the small preference for truth-telling is slightly larger at $t = 1$ than at $t = 2$ (the alternative specification is briefly discussed in Appendix C).[11] Formally, for some ε_1, ε_2 with $\varepsilon_1 > \varepsilon_2$ assumed to be small, party I's payoff as a function of (s, a, m_1, m_2) is:

$$U_I(s, a, m_1, m_2) = u(a, s) - \varepsilon_1 1_{m_1 \neq s} - \varepsilon_2 1_{m_2 \neq s}. \tag{2}$$

[11] The bigger concern about telling the truth at $t = 1$ can be motivated whenever the first message is (slightly) more likely than the second message to be made public, thereby making a lie at $t = 1$ a bit more costly than a lie at $t = 2$ for reputational reasons.

Second, I will be assuming that were party U to receive twice a message $m_1 = m_2 = s \in S$ that would never been sent in equilibrium by party I, party U would make the inference that the state is s. Formally, this can be rationalized by assuming that with probability ε (again assumed to be small), party I in every state s tells the truth twice $m_1 = m_2 = s$ while optimizing on his communication strategy otherwise, i.e. with probability $1 - \varepsilon$.

Third, I will be assuming that when faced with the same belief and the same preference over a subset of messages, the strategy restricted to this subset is the same. This assumption can be rationalized if one considers (small) stochastic perturbations to the payoffs with state-independent distributions (for example as dictated by some exogenous (small) stochastic preference for the various possible messages).

In the next Sections, I characterize the equilibria with forgetful liars of the above perturbed communication game for the limiting case in which ε tends to 0 while keeping ε_1, ε_2 small but fixed. It is worth mentioning that if parties had perfect recall and were rational, one would get equilibrium outcomes corresponding to some equilibrium outcome of the (one-shot) strategic communication game of Crawford and Sobel (1982) in the limit as $\varepsilon, \varepsilon_1, \varepsilon_2$ go to 0.[12] Departures from the standard cheap talk predictions will thus be caused by the imperfect memory of party I.

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.[13] 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)).

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. If instead the forgetful liar remembers his strategy, the knowledge of the state s together with the strategy would lead party I

[12] Mixed strategies are required though.

[13] The main motivation for this is that there is no joint record of the state and the lie in the feedback provided about past interactions.

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

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

whenever $\sum_{m' \in M,\, m' \neq s} \sigma_1(m' \mid s) > 0$.[14] 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). This approach is in the spirit of either of the multiselves approaches to imperfect recall as defined in Piccione and Rubinstein (1997). I would suggest it is a legitimate formulation in environments in which the same economic agent would be in the role of party I many times so that party I would more naturally be viewed as having a good sense of how he routinely behaves (even if not physically remembering which lies he previously made in the current interaction).[15] 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 $s' \neq s$. If instead such a player $I_2(s)$ were aware he made a lie, it would then be natural to assume that he would rule out $m_1 = s$ and a new definition of belief (conditioning the aggregate distribution of lies on $m_1 \neq s$) should be considered.[16] The equilibria characterized below would remain equilibria with this modification, and as discussed later no other equilibria employing pure strategies would

[14] If the probabiity of lie in state s is 0, some trembling is required to pin down the expectation.

[15] 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 (and the main approach assumes that only the marginal distribution of lies is considered instead).

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

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

arise with this alternative approach.

3 Pure persuasion

In this Section, I assume that for all a and s, $u(a,s) = v(a)$ for some increasing function $v(\cdot)$. That is, 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.

A simple class of strategies.

I consider the following family of communication strategies for party I referred to as (s^l, s^h)-communication strategies. Party I in state s sends twice the same message $m_1(s) = m_2(s)$ whatever $s \in S$. There are two types $s^h, s^l \in S$ with $s^h > s^l$ such that all types $s \leq s^l$ lie and say s^h , i.e. $m_1(s) = m_2(s) = s^h$, and all types $s > s^l$ say 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, the induced aggregate distribution of lie at $t = 1$ is a mass point on s^h.

Second, the best-response of party U is to choose $a(s) = s$ for $s \in S$ whenever $m_1 = m_2 = s \neq s^h$ and (approximately as ε goes to 0) $a(s^h) = a^E(s^l, s^h) = E(s \mid s \leq s^l$ or $s = s^h)$ when $m_1 = m_2 = s^h$.[17]

Third, 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 $a^E(s^l, s^h) - \varepsilon_1 - \varepsilon_2 \geq s^l$.

Fourth, if party I with type $s > s^l$ were to lie at time $t = 1$, he would believe at time $t = 2$ that he said $m_1 = s^h$ according to the proposed solution concept. By lying and saying $m_1 = s^h$ at time $t = 1$, party I with type s could ensure to get $a^E(s^l, s^h) - \varepsilon_1 - \varepsilon_2$ just assuming party $I_2(s)$ wants to avoid that inconsistent messages are being sent (it will be shown to be a necessary requirement in equilibria employing pure strategies). Thus,

[17] This follows because the action scheme just defined leads to the true expected value of s after two consistent messages $m_1 = m_2 \in S$ where this expectation for $m_1 = m_2 \leq s^l$ is pinned down by the trembling hand truth-telling assumption.

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 $s^l_+ \geq a^E(s^l, s^h) - \varepsilon_1 - \varepsilon_2$.

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 requests that party I uses an (s^l, s^h)-communication strategy for some (s^l, s^h) satisfying $s^l_+ \geq a^E(s^l, s^h) - \varepsilon_1 - \varepsilon_2 \geq s^l$. 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 (for all s) can be sustained is shown easily by setting $a_{inc} = 0$, fixing party I's belief in case of lie to be that $m_1 = 0$ was sent with probability 1, and requiring that $a(s) = s$ for all $s \in S$.[18] However, such an equilibrium is somehow fragile, as it is not robust to a number of perturbations, for example ones in which party I would exogenously make lies in S other than $m_1 = 0$ with strictly positive probability.[19]

The more interesting aspect of Proposition 1 concerns the characterization of the equilibria with some lying activity. The following observations are the key drivers of it.

First, no matter what the state s is, it cannot be rewarding for party I to be inconsistent as compared with any alternative strategy that would be employed in equilibrium by party I possibly in other states s'. This essentially follows because no matter what the state s is, it is very easy for party I to be inconsistent (first tell the truth, then lie) so that party I, no matter what s is, should be getting at least what he can get by being inconsistent. This in turn paves the way to establishing that the informed party never sends inconsistent messages in a pure strategy equilibrium.[20]

Second, because the belief of party I at $t = 2$ about m_1 is the same whatever s after a lie at $t = 1$, it must be, in a pure strategy equilibrium, that the same choice of m_2 is made

[18] With this in place, after a lie at $t = 1$, party I with type s should optimally send $m_2 = s$ as he would believe he previously sent $m_1 = 0$ and would optimally decide to be truthful at $t = 2$, i.e. $m_2 = s$, so as to avoid the extra penalty ε_2 incurred when sending $m_1 = 0$.

[19] With such perturbations, party I with type $s = 0$ would be strictly better off lying.

[20] To establish this formally, I make use of an unravelling argument focusing on the set of states in which party I would opt for inconsistency (that is, showing by contradiction that if this set were not empty, the highest such type would be strictly better off telling the truth).

after a lie at $t = 1$ given that preferences are state-independent.[21] Call m^* this unique time 2 message sent after a lie at $t = 1$. Clearly, in an attempt to avoid being inconsistent (which is not rewarding as just noted), and anticipating that message m^* will be sent next at $t = 2$ if party I lies at $t = 1$, it is best for party I to send message m^* today at $t = 1$ if not telling the truth. This establishes that in a pure strategy equilibrium, only one lie m^* can be made both at $t = 1$ and 2. This in turn implies given the expectation (1) that, no matter what the state s is, in case party I lies at $t = 1$, the belief of I at $t = 2$ must be that $m_1 = m^*$ was sent at $t = 1$.

Thus, reminding that $a(m^*)$ denotes party U's action after $m_1 = m_2 = m^*$ was sent, party I in state $s \neq m^*$ either engages in making twice the lie m^* expecting to get $a(m^*) - \varepsilon_1 - \varepsilon_2$ 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 to the extent that $s \neq m^*$ would not be a lie made by party I in equilibrium no matter what the state s is).[22]

As a result, party I in state $s \neq m^*$ chooses to make the lie m^* whenever $s < a(m^*) - \varepsilon_1 - \varepsilon_2$, and he tells the truth whenever $s > a(m^*) - \varepsilon_1 - \varepsilon_2$. 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, $a(m^*)$ takes value $a^E(s^l, s^h)$ with $s^h = m^*$ and s^l being such that $s^l_+ \geq a^E(s^l, s^h) - \varepsilon_1 - \varepsilon_2 \geq s^l$.

The detailed formal derivation of Proposition 1 appears in Appendix A. Here, I provide several additional comments that can help understand better some properties of the approach. 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 $s^l_+ > s_2 \geq a^E(s^l, s^h) - \varepsilon_1 - \varepsilon_2 \geq s^l$ when ε_1 and ε_2 are small enough. More generally, the pure strategy equilibria with some lying activity 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

[21] This makes use of the assumption that perturbations are state-independent too so as to deal with indifferences. In the detailed proof shown in Appendix, one has to worry about the state-dependence induced by the slight preference for truth-telling, but the same conclusion goes through.

[22] This is making use of the truth-telling perurbation assumption in case in state s party I would choose not to tell the truth.

follows the (s^l, s^h)- communication strategy.[23]

Second, I note some close analogy between the shape of a pure strategy equilibrium with lie $m^* \in S \setminus \{0\}$ and the Perfect Bayes 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 $s = m^*$.[24] 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 saying m^* at $t = 2$ (so as to avoid being inconsistent). 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). This analogy is the consequence of the observation that no inconsistent messages are sent in a pure strategy equilibrium with forgetful liars and when there is some lying activity there is exactly one lie being made in such an equilibrium.

While Proposition 1 only considers equilibria employing pure strategies, I show that equilibria employing mixed strategies can also arise. These are characterized in Appendix B. In mixed strategy equilibria, multiple lies are chosen in a probabilistic way whenever party I is of a type below some endogenously determined threshold (denoted a^* in appendix B); when lying, the distribution over lies is the same at $t = 1$ and 2 and the same for all lying types;[25] sometimes inconsistent messages are sent (due to the independent randomization at $t = 1$ and 2), and inconsistent messages result in an action a_{inc} that is strictly below the one arising when the same lie (in the support of the distribution of lies) is being made at $t = 1$ and 2.

Assuming that forgetful liars remember than they lied would not affect the equilibria shown in Proposition 1 nor the equilibria in mixed strategies analyzed in Appendix B because if in state s party I lies, it is never the case in any of these equilibria that

[23] Observe that for all $s^h \in S \setminus \{0\}$ we have that $a(0, s^h) > \varepsilon_1 + \varepsilon_2$ for ε_1, ε_2 small enough, and $a(s^h, s^h) - s^h < 0$. Thus, choosing s^l to be $\max_s\{s \text{ such } a(s, s^h) > s + \varepsilon_1 + \varepsilon_2\}$ guarantees that all conditions are satisfied.

[24] In the continuous type space, 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.

[25] This aspect follows because I am considering state-independent perturbations.

s is a lie made in equilibrium by party I in another state $s' \neq s$. As it turns out, it can be shown that no other equilibrium employing pure strategies would arise with this alternative approach (see Appendix C). As a further robustness check, I also briefly consider in Appendix C the case in which $\varepsilon_2 > \varepsilon_1$ (so that party I has a stronger preference for truth-telling at $t = 2$ than at $t = 1$), and the case in which party U would observe the two messages (m_1, m_2) fully no matter whether they are consistent or not. In both cases, I note that the same equilibria in pure strategies as in Proposition 1 arise.

3.1 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 are close to the truth-telling equilibrium, resulting in the approximate first-best outcome for party U. More precisely,

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'.*

Proposition 2 *Consider a sequence $(S^n)_{n=\underline{n}}^{\infty}$ of state spaces such that, for each n, S^n satisfies the n-fine grid property. Consider a sequence $(\sigma^n)_{n=\underline{n}}^{\infty}$ of equilibria with forgetful liars associated with S^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 for pure strategy equilibria and of Proposition 4 (see Appendix B) for mixed strategy equilibria. Let a^* be the expected payoff obtained by party I when engaging in a lie at $t = 1$ (in an equilibrium either in pure or in mixed strategies). 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 s that either lie below a^* or else $s = 1$ must be significantly below a^* (at a distance at least $\frac{2}{n}$) 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 detailed argument appears in Appendix D.

Restricting attention to pure strategy equilibria, the intuition for Proposition 2 can be understood as follows. As discussed around Proposition 1, an equilibrium with forgetful liars in pure strategy with lie m^* can be viewed as a Perfect Bayes 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, one gets an equilibrium outcome close to that in the classic persuasion game in which the unravelling argument leads to full disclosure. The $n-$fine grid property precisely ensures that the ex ante probability of any type $s \in S$ gets very small in the limit as n gets large, thereby explaining the limit first-best result.

3.2 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 can still arise with this alternative approach, the main observation is that many additional equilibrium outcomes can also be sustained. 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 and says $m_t = \overline{s}_k$ 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 $a(m_1 = m_2 = \overline{s}_k) = E(s \in S_k)$, and one may require, for example, that $a_{inc} = 0$ so that party I whatever his type is not tempted to send inconsistent messages, and also that the belief of $I_2(\overline{s}_k)$ if $I_1(\overline{s}_k)$ were to lie is that message 0 was sent at $t = 1$.

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 $I_1(\underline{s}_k)$ said $m_1 = \overline{s}_k$ 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 say $m_2 = \overline{s}_k$ as any other message is perceived to trigger a_{inc}, and $a_{inc} = 0 < E(s \in S_k)$. 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 says $m_1 = \overline{s}_k$ or else he tells the truth. Given that $E(s \in S_k) > \underline{s}_k$, he strictly prefers lying (whenever ε_1, ε_2 are small enough, i.e. $\varepsilon_1 + \varepsilon_2 < E(s \in S_k) - \underline{s}_k$), 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 obtained using the off-path beliefs proposed above.[26]

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 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 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).

[26] 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 perturbation, 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 say $m_1 = 1$ anticipating that $I_2(\overline{s}_k)$ would say $m_2 = 1$ next, and this would be worse than truth-telling.

4 Communicating with state-dependent objectives

I consider now alternative specifications of party I's preferences in which I's blisspoint action may depend on the state. Specifically, $u(a,s) = -(a-b(s))^2$ 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 with such state-dependent objectives, multiple lies may arise in equilibrium. 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 equilibria with forgetful liars employing pure strategies, 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 $b(s) = s+\beta$ for some β satisfying $\frac{1}{2} > \beta > 0$.

I will look for conditions on s_1^*, s_2^* so that in states $s=0$ and s_1^*, party I says he is of type s_1^* at times $t=1$ and 2 (i.e., party I in those states sends the messages $m_1 = m_2 = s_1^*$), and in states $s=s_2^*$ and 1 party I says he is of type 1 at times $t=1$ and 2 (i.e., he sends the messages $m_1 = m_2 = 1$). I will impose that in case of inconsistent messages ($m_1 \neq m_2$), party U chooses $a_{inc} = 0$ (which can be rationalized by requiring that with some small probability, party I when of type $s=0$ sends messages at random at $t=1$ and 2).

In such a scenario, party U must choose $a(m_1 = m_2 = s_1^*) = \frac{s_1^*}{2}$, $a(m_1 = m_2 = 1) = \frac{s_2^*+1}{2}$ and $a(0) = 0$, $a(s_2^*) = s_2^*$. Moreover, two lies $m_1^* = s_1^*$ and $m_2^* = 1$ are made in equilibrium, and these two lies are overall equally likely. Thus, party I in state s after a

lie $m_1 \neq s$ at $t = 1$ believes at $t = 2$ that at $t = 1$ he either sent $m_1 = s_1^*$ or 1 each with probability half.

To be an equilibrium, it should be that party I in state $s = s_1^*$ weakly prefers $a(s_1^*)$ to $a(1)$, as otherwise, party I would strictly prefer saying $m_1 = 1$ at $t = 1$ anticipating that he would stick to his lie at $t = 2$. That is, $s_1^* + \beta - a(s_1^*) \leq a(1) - s_1^* - \beta$ or

$$\frac{1 + s_1^* + s_2^*}{2} - 2\beta \geq 2s_1^*. \tag{4}$$

Also, it should be 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, $a(1) - s_2^* - \beta \leq s_2^* + \beta - a(s_1^*)$ or

$$2s_2^* \geq \frac{1 + s_1^* + s_2^*}{2} - 2\beta. \tag{5}$$

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, $a(s_1^*) < 2\beta$ or

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

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, $a(1) < s_2^* + 2\beta$ or

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

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 $\beta < \frac{1}{2}$), the above two-lie communication strategy can be sustained as an equilibrium with forgetful liars. ♣

Characterization of equilibria employing pure strategies.

To provide a simple characterization, let me assume that for any two distinct pairs (N_1, N_1') and (N_2, N_2') such that N_1, N_1', N_2, N_2' are subsets of $N = \{1, ..., n\}$, it is not the case that $p(N_1)E(s_k,\, k \in N_1') = p(N_2)E(s_k,\, k \in N_2')$ where $p(N_i)$ denotes the sum of p_k for $k \in N_i$ (such a condition is satisfied generically). Let me also perturb the description of

the communication game as defined in Section 2 by assuming that with a tiny probability ε_0, party I with type $s = 0$ randomizes over all possible messages in an independent way at $t = 1$ and 2.[27] The other perturbations parameterized by ε, ε_1, ε_2 ($\varepsilon_1 > \varepsilon_2$) are maintained, and I will be concerned with describing the set of pure strategy equilibria in the limit in which ε, ε_0, ε_1, ε_2 (with $\varepsilon_1 > \varepsilon_2$) as well as $\varepsilon_0/\varepsilon$ go to 0.

Roughly, such equilibria satisfy the following properties. No inconsistent messages are sent in equilibrium, thereby implying (because of the ε_0 perturbation) that $a_{inc} = 0$. 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 $L_k^- = L_k \setminus \{\max(s \in L_k)\}$ and $L = (L_k)_k$. Let $\widehat{a}_k(L) = E(s \in L_k)$ and $(\widehat{p}_k(L))_k$ 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 which is proven in Appendix E summarizes the main properties of the pure strategy equilibria with forgetful liars.

Proposition 3 *There always exists an equilibrium with forgetful liars in pure strategies and any 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 saying $m_1 = m_2 = m_k^*$; 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 inconsistent messages chooses $a_{inc} = 0$; when hearing $m_1 = m_2 = m_k^*$ chooses $a = \widehat{a}_k(L)$; and when hearing $m_1 = m_2 = s \in S \setminus \{m_1^*, ...m_K^*\}$ chooses $a = s$.*

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 and say $m_{k(s)}^*$ 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.

[27] While this allows me to pin down the equilibrium value of a_{inc}, the action chosen by party U when inconsistent messages are being sent, no essential qualitative features of the equilibria shown below depend on this extra perturbation.

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 say a lie m_k^*).[28] 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 3, 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 $b(s) \geq s + \beta$ for some $\beta > 0$, 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.

5 Discussion

5.1 Back to criminal investigations

As highlighted throughout the paper, 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

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

a lying suspect pretends he is not guilty (i.e., by saying $m_1 = 1$ at $t = 1$) he may well remember at $t = 2$ that he previously said so, making the memory asymmetry assumption as considered in the main model not so clearly compelling in this case.

Yet, in the criminal investigation application (as well as in many other applications), the full description of the state (or event) typically consists of many more details that just the level of guilt of the suspect. I wish now to explore in the pure persuasion context a setting in which a lying suspect would not remember the details he reported when lying (by contrast, a suspect telling the truth would remember those details).

There are obviously many ways of modeling this. I will propose a simple one that I think captures the essential ingredients that are relevant for this application. The main insight will be that if the protocol through which the suspect is requested to provide the details of the event (or state) is sufficiently non-straightforward, a similar analysis as the one obtained in the main model arises (in particular, at most one lie will be shown to arise in equilibrium, and as the grid of the possible levels of guilt gets finer and finer, one approximates the first-best in which the level of guilt is elicited almost for free). By contrast, if the protocol is too simple (such as requesting to provide the details always in the same frame), the suspect is able to engage in quite effective lying activity and the full revelation of the state may not be taken for granted (even in the fine grid case).

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 of the state θ, and $s = s_A + s_B$ summarizes the characteristics of the state (guilt level) parties I and U care about. Specifically, 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 $p(s_k)$ as in the main model.

The communication protocol -that should be thought of as non-straightforward- takes the following form. At $t = 1$, party I is requested 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 requested 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. 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 $\widehat{s}$. If the two messages are inconsistent, then party U is informed of the inconsistency. To simplify the exposition, I am assuming that in case of inconsistent messages, party U chooses an action that is very detrimental to party I (say a_{inc} is set sufficiently low-this can be endogenized as in the main model).

At $t = 2$, party I in state $\theta = (s_A, s_B)$ has perfect memory of m_1 if he told the truth at $t = 1$, i.e. if he said $\overrightarrow{m_1} = (s_A, s_B)$ when asked to describe the state in normal order or $\overleftarrow{m_1} = (s_B, s_A)$ when asked to describe the state in reverse order.

If however at $t = 1$, party I lied, then at $t = 2$, party I 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) reported in m_1 when there was a lie at $t = 1$.[29] In all cases, party I remembers the state $\theta = (s_A, s_B)$.

I am also perturbing the above specification, as in the main model, by assuming that party I has a slight preference for truth-telling, and that with small probability, party I engages in truth-telling without thinking about it (and also that faced with the same belief after a lie at $t = 1$, party I uses the same strategy irrespective of the state θ and of $X = A$ or B).

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

I will now sketch here the main arguments why the pure strategy equilibria with forgetful liars in this extended 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 more straightforward- 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 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

[29] 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 $\overleftarrow{m_1} = (\widehat{s}_B, \widehat{s}_A) \neq (s_B, s_A)$.

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 said $\widehat{s}_X = \frac{s^h}{2}$ at $t = 1$ whether $X = A$ or B. To avoid being inconsistent, he will choose to say $m_2 = \frac{s^h}{2}$ 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 engages in 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), and either for $X = A$ or B, party I will be reported to have been inconsistent. This in turn (given the assumption that being inconsistent is very detrimental) deters party I from engaging in 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 in this case, the support of the equilibrium distribution of $\widehat{s}_X$ conditional on a lie being made must 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 make the same report of m_2 whether requested to report attribute $X = A$ or B (this makes use of the assumption that a player faced with the same belief and the same preferences should be choosing the same strategy). As a result, party I for at least one lie and one realization of X would be reported as being inconsistent. Party I would prefer avoiding this (if inconsistency is sufficiently harmful) by being truthful throughout, thereby explaining why it is not possible to support equilibria with multiple lies in this extended 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 $m_1 = (\frac{\widehat{s}}{2}, \frac{\widehat{s}}{2})$ at state θ, party I can ensure not being inconsistent in such a variant). As in Subsection 3.2, 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 $= 2$ is always asked to report the A attribute (instead of randomizing between attributes A and B), one could again support many more lies in the pure strategy equilibria with forgetful liars (sticking to the same modeling of the expectation of a forgetful liar as in the main model). 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) who advocate in favor of the use non-trivial frames when asking multiple questions to a suspect.[30]

5.2 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 considered in Subsection 3.2 (or in Piccione and Rubinstein (1997)) in which party I would be assumed to know how his lying strategy varies with the state. The second concerns whether if viewing party U as committing to some pre-specified course of action (as a function of the outcome of the communication) leads to the same equilibrium analysis

[30] 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.

as in the main model in which there is no such commitment.

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 said $m_1 = s_1^*$ 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 said $m_1 = 1$ 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 report $m_2 = 1$ 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 deeper consequences if we modify the 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 he said at $t = 1$ he attaches a (subjective) probability 0.5

that he may be declared inconsistent (resulting in $a_{inc} = 0$) 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 said $m_1 = 1$ 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 $a(m)$ in case $m_1 = m_2 = m$ and choosing a_{inc} in case $m_1 \neq m_2$ (while party I 's memory problems would still 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 $a(m)$ and a_{inc} are set as in the corresponding equilibrium. A more interesting observation is that fixing $a(m)$ and a_{inc} 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 $a(m)$ and a_{inc} are fixed, 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 is suggestive that from a full implementation perspective, there may be a potential benefit in the absence of commitment in environments with forgetful liars.[31]

To see this, consider the pure persuasion case, and assume that $a(1)$ is set close to 1 while a_{inc} is set at a low level (say 0), and $a(m)$ is set below 1 for all $m < 1$ (which 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 $a(1) = 1$ 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 says $m_1 = 1$ at $t = 1$, he can safely anticipate

[31] 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.

he will choose $m_2 = 1$ at $t = 2$ (so as to avoid being inconsistent). This strategy results in action $a(1) = 1$, and this strategy is optimal given that $a(1) = 1$ is larger than a_{inc} (that would result if party I were to make another lie at $t = 1$) and $a(s)$ 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 $a(1) = E(s)$ in the context of the main model while now she is committed to choosing $a(1) = 1$) 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.[32]

5.3 Inconsistencies and richer memory settings

Partial memory of lies

Sticking to the description of the state s as a single-valued parameter as in the main model (and unlike what discussed in subsection 5.1), one can consider memory settings in which party I when lying at $t = 1$ would have at $t = 2$ a partial memory of m_1 (instead of no memory at all). This could be modeled by assuming that in case of lie, party I at $t = 2$ receives a noisy signal η about m_1. Party I at time $t = 2$ after a lie at $t = 1$ would then form an updated belief about m_1 taking both into account the signal η and the aggregate distribution of lies as defined in Section 2 (serving here the role of the prior). In such a scenario, if there are several lies being made in equilibrium, then it may well be for some signals η that after a lie m_1^* at $t = 1$, party I is led at time $t = 2$ to believe that he is more likely to have sent another lie m_2^*. If one considers a setting in which party U would choose a very detrimental action in case of inconsistency, this would lead party I at time $t = 2$ to choose $m_2 = m_2^*$ then, which would result for party I in a poor expected consequence of engaging into the lie m_1^* at time $t = 1$. Under natural specifications of the signal structure, such considerations imply that it would not be possible to support

[32] A similar observation does not arise in the classic certification environment. There, if all types can be certified and party U commits to $a = 0$ if the state is not disclosed, only the first-best arises, exactly as in the game without commitment (a similar comment applies to Dye's setting). Observe that I have taken commitment to mean that within the communcation game, party U would commit to $a(m)$ and a_{inc} ex ante before the actual communication takes place. Of course, if party U were free to choose whatever mechanism and commit to it, she would be able to do at least as well as in the communication game of Section 2 simply by choosing not to commit and play this game.

multiple lies in equilibrium even when party I's objective is state-dependent as in Section 4. Clearly, the equilibria with only one lie obtained in the main analysis are unaffected by the possibility of partial memory of liars, as in the one-lie case, the signal η is not needed to know what the lie was. As an alternative to imposing large punishments in case of inconsistency, one may consider having more rounds of communication so that party U can make use of the richness of the history of messages to detect lies. A careful analysis of such a scenario is left for future research.

Inconsistencies in richer contexts

In the above setting, I have assumed that the information concerning s held by party I did not change over time. In some applications, it may be natural to consider cases in which party I could either learn more about the state with time or forget some aspects of the state as time elapses. For example, suppose party I is a witness of a potential crime scene. It is unfortunately common that a witness may not remember all details equally well, as a number of details may be less salient to a witness than to a suspect. Also, with well designed cues, such a witness having forgotten some less salient details may be led to recover those. Such extensions with possibly changing information on the state s would deserve further research, but one can already indicate that any of these extensions would call for considering more nuanced notions of inconsistency, for example identifying two messages m_1 and m_2 at times $t = 1, 2$ as inconsistent only if it would not be possible to explain them through a change of I's information (or memory) about the state s.

Lie by omission and memory

So far, I have treated any message other than the full truth (as known by party I) as a lie. But, one could make a distinction between false statements (that would be incompatible with the truth according to accepted meaning) and lies by omission in which party I would withhold part of the truth. One could reasonably argue that a party when not telling the whole truth would remember the type of lie he made (by omission or otherwise). In the spirit of the above modeling, one could then assume that when a lie by omission was made, party I would be aware that no false statement was made, even if not remembering which aspect of the truth was communicated. A precise modeling of this (in particular dealing with the aggregation over states) would deserve further work.

Appendix A (Proof of Proposition 1)

I first establish a few results that apply to all equilibria with some lying activity whether in pure or in mixed strategies.

Lemma 1 *Suppose $I_1(s)$ tells the truth. Then party I with type s gets* $\max(a(m_1 = m_2 = s), a_{inc} - \varepsilon_2)$.

Proof. After $m_1 = s$, $I_2(s)$ would choose $m_2 = s$ if $a(m_1 = m_2 = s) \geq a_{inc} - \varepsilon_2$ and $m_2 \neq s$ otherwise yielding the result. ♣

Lemma 2 *Suppose m is a lie made with positive probability at time $t = 1$ by some $I_1(s)$, $s \neq m$. Then $a(m_1 = m_2 = m) \geq a_{inc} + \varepsilon_1$.*

Proof. Suppose by contradiction that $a(m) < a_{inc} + \varepsilon_1$ and $m_1(s) = m \neq s$. By saying $m_1 = m$, $I(s)$ gets $\max(a(m) - \varepsilon_1 - \varepsilon_2, a_{inc} - \varepsilon_1)$. By saying $m_1 = s$, $I(s)$ gets $\max(a(m_1 = m_2 = s), a_{inc} - \varepsilon_2)$ (see lemma 1), which is strictly larger than $a_{inc} - \varepsilon_1$ because $\varepsilon_1 > \varepsilon_2$. Thus, $I_1(s)$ cannot choose $m_1 = m$ providing the desired result. ♣

Lemma 3 *If $I_1(s)$ says $m_1 = m \neq s$ with strictly positive probability, it must be that $I_2(s)$ says $m_2 = m$ with strictly positive probability.*

Proof. If $I_2(s)$ does not choose $m_2 = m$, $a_{inc} - \varepsilon_1$ would be obtained at best by $I_1(s)$, and by lemma 1, $I_1(s)$ would be strictly better off saying $m_1 = s$. ♣

The next lemma is specific to equilibria employing pure strategies.

Lemma 4 *In an equilibrium with forgetful liars employing pure strategies, there can be at most one lie at $t = 1$.*

Proof. Suppose $I(s)$ lies and says $m_1 = m \neq s$ and $I(s')$ lies and says $m_1 = m' \neq s'$. By lemma 3, the same lie must be repeated at $t = 2$. $I(s)$ by saying $m_1 = m_2 = m$ gets $a(m) - \varepsilon_1 - \varepsilon_2$. If $m' \neq s$ and $I_1(s)$ says $m_1 = m'$, he must find $m_2 = m'$ optimal (as $I_2(s')$ finds $m_2 = m'$ optimal). Thus, he must pick $m_2 = m'$ (as does $I_2(s')$)[33] so that one

[33] This makes use of the requirement that $I_2(s)$ and $I_2(s')$ having the same preferences over $m_2 \neq s, s'$ should choose the same best-response.

should have $a(m) - \varepsilon_1 - \varepsilon_2 \geq a(m') - \varepsilon_1 - \varepsilon_2$. If $m' = s$ and $I_1(s)$ says $m_1 = m'$, then $I(s)$ gets at least $a(m')$. Thus, in all cases, $a(m) \geq a(m')$. By a symmetric argument, one should also have $a(m') \geq a(m)$, and thus $a(m) = a(m')$. I next observe that it cannot be that m' is equal to s as otherwise, $I(s)$ would strictly prefer telling the truth rather than saying $m_1 = m$. Thus, m and m' are both different from s and s' and $I_2(s)$ and $I_2(s')$ should thus pick the same m_2,[34] leading to a contradiction (since $I_2(s)$ should be saying m and $I_2(s')$ should be saying m'). ♣

Following lemma 4, I let m^* denote the unique lie made in an equilibrium with forgetful liars employing pure strategies. The next lemma establishes that inconsistency cannot arise in a pure strategy equilibrium.

Lemma 5 *There can be no inconsistent messages in equilibria employing pure strategies.*

Proof. Assume by contradiction that inconsistent messages can be sent in an equilibrium in pure strategy and call $S_{inc} = \{s \in S \text{ such that } m_1(s) \neq m_2(s)\}$. One should have $a_{inc} = E(s \in S_{inc})$ by the optimality of party U' strategy. Because lies are costly and more so at $t = 1$, if $s \in S_{inc}$, one should have $m_1(s) = s$. Moreover by lemma 2, $a(m^*) \geq a_{inc} + \varepsilon_1$, and thus, if $m^* \in S$, party I with type m^* after the truth being told at $t = 1$ would strictly prefer telling the truth at $t = 2$, thereby implying that $m^* \notin S_{inc}$. We thus have that $m^* \notin S_{inc}$ (whether or not $m^* \in S$). Consider $s_{inc}^{\max} = \max S_{inc}$. By telling the truth twice, $I(s_{inc}^{\max})$ gets $s_{inc}^{\max}$ (since $s_{inc}^{\max} \neq m^*$ and thus $a(s_{inc}^{\max})$ is pinned down by the truth-telling trembling behavior of $I(s_{inc}^{\max})$). Since $\max S_{inc} > E(s \in S_{inc}) - \varepsilon_2$, it must be that $I(s_{inc}^{\max})$ strictly prefers telling the truth, thereby implying the absurd conclusion $s_{inc}^{\max} \notin S_{inc}$. ♣

By lemma 2, it should be that $a(m^*)$ satisfies $a(m^*) \geq a_{inc} + \varepsilon_1$. Given that there is one lie m^*, the belief of $I_2(s)$ if $I_1(s)$ lies must be that $m_1 = m^*$ was sent with probability 1 at $t = 1$. Given that $a(m^*) \geq a_{inc} + \varepsilon_1 > a_{inc} + \varepsilon_2$, $I_2(s)$ would then find it strictly optimal to say $m_2 = m^*$. Given the expectation that when $I_1(s)$ lies, $I_2(s)$ says $m_2 = m^*$ and given that $a_{inc} < a(m^*)$, $I_1(s)$ if he lies, says $m_1 = m^*$. So for any s, either $I_1(s)$ tells the truth $m_1 = s$ expecting to get $\max(a(m_1 = m_2 = s), a_{inc} - \varepsilon_2)$ by lemma 1 or lies and says $m_1 = m^*$ expecting to get $a(m^*) - \varepsilon_1 - \varepsilon_2$.

[34] This is again using the assumption that with the same preferences and the same beliefs, choices should be the same.

To sum up, for any $s \neq m^*$, the choice of $I(s)$ is between truth-telling resulting in $a(s) = s$ (because s is not in the support of equilibrium lie) or lying twice according to $m_1 = m_2 = m^*$ resulting in $a(m^*) - \varepsilon_1 - \varepsilon_2$ where $a(m^*) = E(s \in S^*)$ with $S^* = \{s \in S \text{ such that } I_1(s) \text{ says } m_1 = m^*\}$ (using the best-response of party U). And $I(m^*)$ can do no better than telling the truth (as a lie at $t = 1$ would result in $a_{inc} - \varepsilon_1$ and $a_{inc} - \varepsilon_1 \leq a(m^*)$). I let s^* denote $\max\{s \in S^*\}$.

If S^* consists only of s^* then $s^* = 0$ as otherwise any $s < s^* = E(s \in S^*)$ would strictly prefer to lie as does s^* contradicting the premise that S^* is a singleton. But when $s^* = 0$, party I with type s^* would strictly prefer telling the truth due to the $-\varepsilon_1 1_{m_1 \neq s^*} - \varepsilon_2 1_{m_2 \neq s^*}$ terms, violating the premise that $I(s^*)$ is lying.

If S^* contains at least two s_k and if $s^* \neq m^*$, then party I with type s^* would strictly prefer $a = s^*$ to $a(m^*) = E(s \in S^*)$ leading him to tell the truth rather m^* at $t = 1$ in contradiction with the equilibrium assumption.

Thus, it must be that S^* contains at least two states and that the lie m^* is the maximal element s^*. The requirement that for $s \neq s^*$, $I(s)$ lies and says m^* whenever $a(m^*) - \varepsilon_1 - \varepsilon_2 > s$ ensures that S^* takes the form $\{s \in S,\, s \leq s_*\} \cup \{s^*\}$ for some s_* with s_* being the largest s in S such that $s < E(s \in S^*) - \varepsilon_1 - \varepsilon_2$. This precisely corresponds to the (s_*, s^*)-communication strategy with $s_*^+ \geq a^E(s_*, s^*) - \varepsilon_1 - \varepsilon_2 \geq s_*$.

That all such communication strategies can be made part of an equilibrium with forgetful liars is easily shown by setting $a_{inc} = 0$, thereby completing the proof of Proposition 1. ♣

Appendix B (Mixed strategy equilibria)

Suppose several lies m_k^*, $k = 1, ..K$, are made in an equilibrium employing mixed strategies. Let a_k denote $a(m_k^*)$ and let μ_k be the probability with which m_k^* is sent at $t = 1$ conditional on a lie being sent then $(m_1 \neq s)$. By lemma 3 it should be that when $I_1(s)$ lied at $t = 1$ and said m_k^*, $I_2(s)$ finds it optimal to say m_k^*. This implies that:

Lemma 6 *Party I with type m_k^* does not lie.*

Proof. Assume by contradiction that $I_1(s)$ lies and says m_k^* with positive probability, $I_1(s')$ lies and says $m_{k'}^*$ with positive probability and $m_{k'}^* = s$. For $I_2(s')$ to find it optimal

to say $m_2 = m^*_{k'}$, one should have $\mu_{k'}a_{k'} + (1-\mu_{k'})a_{inc} \geq \mu_k a_k + (1-\mu_k)a_{inc}$. But, then $I_2(s)$ would strictly prefer saying $m_2 = m^*_{k'}$ so as to save the $\varepsilon_2 1_{m_2 \neq s}$ obtained when $m_2 = m^*_k$. As a result, $I_2(s)$ would never find it optimal to say $m_2 = m^*_k$, violating lemma 3. ♣

The optimality to repeat the same lie (lemma 3) also imposes that $\mu_k a_k + (1-\mu_k)a_{inc}$ be the same for all k. Let a^* denote this constant. Let also denote by μ^2_k the common probability of saying m^*_k at $t=2$ when a lie was made at $t=1$.[35] Given that all lies m^*_k must be chosen at $t=1$ with positive probability, this imposes that $\mu^2_k a_k + (1-\mu^2_k)a_{inc}$ should be the same for all k, which together with the constraint that $\sum_k \mu_k = \sum_k \mu^2_k = 1$ imposes that $\mu^2_k = \mu_k$ for all k.

It is readily verified that $m^*_k \in S$ as otherwise party I with the maximum type $\overline{s}^*_k$ among those who send $m_1 = m^*_k$ with positive probability at $t=1$ would strictly prefer telling the truth (this makes use of $a_k \leq \overline{s}^*_k$).

Moreover, take any s other than m^*_k for $k = 1,..K$. If $s < a^* - \varepsilon_1 - \varepsilon_2$, $I_1(s)$ would strictly prefer saying any m^*_k expecting to get $a^* - \varepsilon_1 - \varepsilon_2$ rather than the truth that would only yield s. If $s > a^* - \varepsilon_1 - \varepsilon_2$, $I_1(s)$ would strictly prefer telling the truth rather than lying. On the other hand, any $I_1(m^*_k)$ would go for telling the truth (using the ε_1 preference for truth telling at $t=1$ and the observation that a lie would not induce a higher expected action). Moreover, for all k, one must have $m^*_k > a^*$ as otherwise lemma 6 would be violated. The above observations imply:

Proposition 4 *Any mixed strategy equilibrium with forgetful liars takes the following form. For some* a^*, $m^*_k > a^*$, $k = 1...K$, *and* $\mu_k > 0$, *with* $\sum_k \mu_k = 1$
$\mu_k a_k + (1-\mu_k)a_{inc} = a^*$
$a_{inc} = E(s \in S, s < a^*)$
$a_k = \frac{\mu_k \Pr(s \in S, s < a^*) a_{inc} + p(m^*_k) m^*_k}{\mu_k \Pr(s \in S, s < a^*) + 1}$
$a(m_1 = m_2 = s) = s$ *for* $s \in S$, $s \neq m^*_k$, $k = 1,...K$.
$I_t(s)$ *with* $s < a^* - \varepsilon_1 - \varepsilon_2$ *says* m^*_k *with probability* μ_k *for* $t = 1, 2$
$I_t(s)$ *with* $s > a^* - \varepsilon_1 - \varepsilon_2$ *says the truth at* $t = 1, 2$.

Observe that when $K = 1$, the conditions shown in Proposition 4 boil down to those in

[35] That μ^2_k is common follows from the requirement that with identical preferences and identical beliefs, the strategy should be the same.

Proposition 1. Moreover, unlike for the equilibria in pure strategies, there are inconsistent messages being sent in mixed strategy equilibria explaining why a_{inc} is pinned down in such equilibria.

Appendix C (Robustness checks)

Pure strategy equilibria when liars remember that they lied

To show that there is no other pure strategy equilibria than the ones shown in Proposition 1 when liars remember that they lied, it suffices to establish that one cannot have an equilibrium in which, for some $s, s' \neq s$ and s", I in state s lies and says $m_s = s$" while I in state s' lies and says $m_{s'} = s$ (as when the type s of a liar is not the lie of some other type, whether or not I in state s remembers that he lied at $t = 2$ does not affect his expectation about which lie he previously made at $t = 1$).

To establish the impossibility, note that the choice of I in state s implies that $a(s") - \varepsilon_1 - \varepsilon_2 \geq a(s)$ (given that I in state s can guarantee $a(s)$ by telling the truth).

If $s" \neq s'$, party I in state s' should then strictly prefer saying $m = s$" rather than $m = s$ since $a(s") > a(s)$ contradicting the premise that in state s', party I should be saying $m = s$. And a fortiori so if $s" \neq s'$ since $a(s") > a(s) - \varepsilon_1 - \varepsilon_2$. ♣

Pure strategy equilibria in pure persuasion games when $\varepsilon_1 < \varepsilon_2$

Consider a pure strategy equilibrium. Let $S_{inc} = \{s \text{ such that } m_1(s) \neq m_2(s)\}$ and $s^*_{inc} = \max S_{inc}$. The main issue is to show that $S_{inc} = \varnothing$ from which it is easy to proceed to show that the equilibria in pure strategy when $\varepsilon_1 < \varepsilon_2$ are the same as those shown in Proposition 1 (when $\varepsilon_1 > \varepsilon_2$). This is established using the observation that if a type engages into inconsistent messages he should first lie and then tell the truth as well as the next lemma

Lemma 7 *s^*_{inc} cannot be a consistent lie, i.e. there is no $s \neq s^*_{inc}$ such that $I(s)$ sends $m_1 = m_2 = s^*_{inc}$.*

Proof. Suppose that $I(s)$ with $s \neq s^*_{inc}$ sends $m_1 = m_2 = s^*_{inc}$. One should have

$$a(s^*_{inc}) - \varepsilon_1 - \varepsilon_2 \geq a_{inc} - \varepsilon_1$$

for $I(s)$ not to prefer sending inconsistent messages, and

$$a_{inc} - \varepsilon_1 \geq a(s^*_{inc})$$

for $I(s^*_{inc})$ not to prefer telling the truth. These two conditions are incompatible. ♣

The above lemma implies that $a(s^*_{inc}) = s^*_{inc}$. Given that $a_{inc} = E(s \in S_{inc})$, this implies that $a_{inc} - \varepsilon_1 < a(s^*_{inc})$ and thus, $s^*_{inc} \notin S_{inc}$ yielding a contradiction.

When Party U observes (m_1, m_2) even when $m_1 \neq m_2$

Consider the variant in which (m_1, m_2) would be observed by party U, even if $m_1 \neq m_2$. It can be shown that it is not possible that messages (m_1, m_2) with $m_1 \neq m_2$ be sent in a pure strategy equilibrium using an argument similar to that used in lemma 5 (the analog of the set S_{inc} considered in the proof of lemma 5 should now be indexed by (m_1, m_2) but the same conclusion arises for each such set), from which one can conclude that the equilibria in pure strategies have the same structure as the ones shown in Proposition 1.♣

Appendix D (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. The condition $\mu_k a_k + (1-\mu_k)a_{inc} = a^*_n$ with $a_k = \frac{\mu_k \Pr(s \in S, s < a^*_n) a_{inc} + p(m^*_k) m^*_k}{\mu_k \Pr(s \in S, s < a^*_n) + 1}$ and $a_{inc} = E(s \in S_n,\, s < a^*_n)$ in the mixed strategy shown in Proposition 4 cannot be satisfied for every k given that $E(s \in S_n,\, s < a^*_n) < a^*_n - \underline{\alpha}\widehat{a}/2(\underline{\alpha} + \overline{\alpha})$; μ_k must be bounded away from 0 irrespective of n (to ensure that $\mu_k a_k + (1 - \mu_k)a_{inc} = a^*_n$); and when μ_k is bounded away from 0, $\mu_k \Pr(s \in S, s < a^*_n)/p(m^*_k)$ grows arbitrarily large with n so that a_k approaches a_{inc} in the limit. This leads to inconsistent conditions, thereby showing the desired result. ♣

Appendix E (Proof of Proposition 3)

Consider a pure strategy equilibrium with forgetful liars. As for pure persuasion games, $\varepsilon_1 > \varepsilon_2$ guarantees that if party $I(s)$ is to engage into sending inconsistent messages, he

would first tell the truth and then lie. 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 (approximately as ε goes to 0) 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 8 *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 saying twice m_k^*, which gives action $a_k = E(s \in L_k) \leq \overline{s}_k = \max L_k$ (and inflicts an extra $\varepsilon_1 + \varepsilon_2$ 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 8 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 9 *There can be no (voluntary) inconsistent messages sent by any type $s \neq 0$ in equilibrium, which in turn implies that $a_{inc} = 0$.*

Proof. Assume by contradiction that there are voluntary inconsistent messages in equilibrium made by some type $s \neq 0$. As already noted, party I with such a type s would first tell the truth $m_1 = s$ and then lie to $m_2 \neq s$. Consider $\overline{s}_{inc} = \max \{s \text{ such that } m_1(s) \neq m_2(s)\}$. $\overline{s}_{inc}$ is not one of the m_k^* because $\overline{s}_{inc}$ is none of the $\overline{s}_k$ and Corollary 1 holds. It follows that party I with type $\overline{s}_{inc}$ would be strictly better off by telling twice the truth rather than by sending inconsistent messages (this makes use of the perturbation $-\varepsilon_2 1_{m_2 \neq s}$ when there is only one s sending inconsistent messages), thereby leading to a contradiction. That $a_{inc} = 0$ follows then from the perturbation that was assumed on the communication strategy of $s = 0$. ♣

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 10 *For any $k_1 < k_2$, if in equilibrium $I(s)$ makes the consistent lie $m_{k_1}^*$ and $I(s')$ makes the consistent lie $m_{k_2}^*$, 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 say $m_2 = m_{k(s)}^*$ 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 $v(k_2, s_2) > v(k_1, s_2)$.[36]

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.

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 and say $\overline{s}_{k_{j+1}}$ for $j < J$ and $\overline{s}_{k_J}$ would lie and say $\overline{s}_{k_1}$, which would violate the property just established. ♣

To complete the description of equilibria, let $L_k^- = L_k \backslash \{m_k^*\}$ where $m_k^* = \overline{s}_k = \max L_k$; $p(L_k^-)$ denote the probability that $s \in L_k^-$; $\mu_k(L) = \frac{p(L_k^-)}{\sum_r p(L_r^-)}$ 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

[36] 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 $b(s_2) > b(s_1)$.

and say $m^*_{k(s)}$ depending on what he likes best and that by Lemma 10 party I with type $\overline{s}_k = m^*_k$ should prefer telling the truth to lying by saying $m^*_{k(\overline{s}_k)}$, the conditions shown in Proposition 3 follow.

Finally, to show that there exists an equilibrium in pure strategies with some consistent lies, think of having a unique lie set, $K = 1$, and set $L_1 = \{s_1, s_2\}$ 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 for equilibrium are satisfied.

That there can be no equilibrium with no consistent lie follows from the observation that in such a case (due to the ε_0 perturbation of the strategy of party I with type $s = 0$), the support of equilibrium consistent lies would assign equal probability to all messages and player $I_1(s)$ would then strictly prefer lying to the message that corresponds to the type $s_k \in S$ that is closest to $b(s)$ anticipating that player $I_2(s)$ will make the same lie and that party U will choose $a = s_k$ (I am using here that $b(s_k) > \frac{s_k + s_{k+1}}{2}$ to ensure that every type would like to be confused with a higher type if possible and that $\varepsilon_0/\varepsilon$ goes to 0 to ensure that the contribution of type $s = 0$ is negligible). ♣

References

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

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

[3] Ben-Porath E., E. Dekel, and B. Lipman (2019): 'Mechanisms with evidence: Commitment and robustness,' forthcoming *Econometrica*.

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

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

[6] Deneckere R. and S. Severinov (2017): 'Screening, signalling and costly misrepresentation,' mimeo.

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

[8] Dziuda, W. and C. Salas (2018): "Communication with detectable deceit," mimeo.

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

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

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

[12] Glazer J. and A. Rubinstein (2014): 'Complex questionnaires,' *Econometrica* 82, 1529–1541.

[13] Green J. and N. Stokey (2007): 'A two-person game of information transmission,' *Journal of Economic Theory* 135, 90–104

[14] Green J. and J. J Laffont (1986): 'Partially Verifiable Information and Mechanism Design,' *Review of Economic Studies* 53 (3): 447–56.

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

[16] Grossman, S. J., and O. D. Hart. (1980): "Disclosure Laws and Takeover Bids." *Journal of Finance* 35 (2): 323–34.

[17] Hart S., I. Kremer and M. Perry (2017): 'Evidence Games: Truth and Commitment,' *American Economic Review* 107, 690-713.

[18] Hörner, J., X. Mu, and N. Vieille (2017): "Keeping your story straight: Truthtelling and liespotting," mimeo.

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

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

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

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

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

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

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

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

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

[28] Vrij A., S. Mann, R. Fisher, S. Leal, R. Milne and R. Bull (2008): 'Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order,' *Law and Human Behavior* 32, 253-265.

[29] Vrij A., P. A. Granhag, S. Mann and S. Leal (2011): 'Outsmarting the liars: Toward a cognitive lie detection approach,' *Current Directions on Psychological Science* 20(1), 28-32.