# A Supernatural Reputation*

Francisco Silva†

March 12, 2017

**Abstract**

I study how someone can successfully sustain the reputation of having supernatural powers. Rational agents believe that psychics, oracles and fortune tellers have supernatural powers even when they do not, because, in their eyes, the data that would be generated by someone with supernatural powers is the same as the one generated by someone who only pretends to have supernatural powers. Experts have an incentive to pretend to have supernatural powers as this increases the number of people who are willing to pay for their advice. Furthermore, I argue that an expert who claims to have supernatural powers may be better for society than an honest expert.

JEL classification: C73, D82
Keywords: reputation, dynamic cheap talk, experts

*I would like to thank Mustafa Dogan, Selman Erol, Nicolas Figueroa, Ju Hu, Samir Mamadehussene, Joaquin Poblete, Francisco Ruiz-Aliseda and Bernardita Vial for their useful comments.

†Instituto de Economia, Universidad Catolica de Chile, Vicuña Mackenna 4860, Piso 3. Macul, Santiago, Chile. (email: franciscosilva@uc.cl).

# 1 Introduction

Throughout the history of civilization, there have always been a number of institutions that have claimed to have some sort of supernatural power, which enables them to predict the future. It is well documented that, during the Antiquity, oracles were quite important in people's daily lives, and even at present time there is an abundance of fortune tellers, psychics and a number of other divinely inspired institutions. In fact, according to a poll from 2005, in the United States, every 3 out of 4 people have some type of paranormal belief. In particular, 41% of those surveyed believed in extrasensory perception (ESP), 32% believed in ghosts and 31% believed in telepathy.[1] In this paper, I study the reasons why people may have these beliefs in the supernatural.

One obvious reason is that these "experts" truly have supernatural powers. In that case, they are certainly of use for their clients so it is no wonder that they have survived until modern times. Nevertheless, I take the point of view of a skeptic, i.e. I assume that no one has supernatural powers. Oracles do not have any contact with the Gods, there is no information hidden in the cards, there is nothing to be learned from looking at a crystal ball. Is it possible that these institutions survive if they do not have special powers? If so, how? How does the lie that they tell survive empirical observation?

One common argument is that the people who actually believe such supernatural stories are somewhat misguided. In the words of economists, the people who believe in the supernatural do not have rational expectations, or they have some sort of behavioral bias which makes them more susceptible to being misled. I confess I am not fully convinced by this explanation. I find it unlikely that someone who is systematically confronted with data disproving the effectiveness of psychics and oracles would simply choose to ignore it. Whatever bias one might have should be overturned by the sheer amount of evidence available, because, after all, oracles and psychics have been around for centuries. Even in Ancient Greece, the idea that only the ignorant would consult the oracles is not accurate. Take the oracle of Delphi, which was the most important oracle of Ancient Greece and lasted for centuries. The clients of the oracle were often wealthy people, in the top of Greek society, with a proper education and, sometimes, an entourage of scholars. Furthermore, there is anecdotal evidence that some of the clients of the oracle of Delphi would go out of their way to gauge its prophetic power, which seems to confirm the idea that they were not just taking the claim that the oracle was somewhat divine at face value.[2]

My approach is the following. I assume that all agents have rational expectations and that they are able to access an arbitrarily large set of evidence. What this implies is that the only way that psychics and oracles can sustain the lie that they have supernatural powers is to have the evidence back it up. In a way, the lie they tell must be empirically indistinguishable

[1] According to a Gallup poll from June, 2005.

[2] For a thorough study and discussion on the oracle of Delphi, see Fontenrose (1978) among others.

from the truth, so that an agent who observes the data and updates his beliefs about the world using Bayes' rule cannot rule out the existence of the supernatural.

There are two main theories on how oracles, and, in particular, the oracle of Delphi, were able to able to survive as long they did.[3] The first theory says that the advice given by the oracle was ambiguous, so that the oracle was never proved wrong. Even if true, this cannot be the whole story. If oracles only gave advice which did not have any informational value there would be no reason for clients to spend money to hear it. The second theory is that those employed by the oracle would try and obtain information from the client (or his entourage) in order to better inform the advice given by the oracle. The powerful and wealthy clients of the oracle of Delphi would, often times, have to stay for several days in the city of Delphi waiting to be received by the priests. During this time, several of the citizens of Delphi, affiliated with the oracle, would try and learn more about the circumstances of the client and report back to the oracle. Once again, this does not appear to be the whole story because this strategy would, at best, leave the oracle with the same information as the client. Therefore, the advice given by the oracle would not give the client any additional information. But if the oracle is simply telling its clients what they already know, the clients would not want to pay for the privilege.

I put forward a theory that combines these two ideas with the notion that the problems of the different clients are somewhat correlated. The information that the oracle (or the psychic for that matter) obtains from past clients is useful when dealing with future clients. For example, imagine that state $A$ is considering invading state $B$ and asks the oracle for advice. The oracle, through its past dealings with state $B$, might know that a war will be damaging for both states and, as a result, might give the (good) advice of restrain. In a more modern context, when someone consults a psychic to seek advice on whether they should divorce their spouse, the psychic is likely to have dealt with a lot of similar situations in the past and should have garnered a considerable amount of expertise on the subject. In a way, oracles and psychics are like experts, who have had access to a lot of information given by their clients, and, as such, are capable of providing sound advice.

While this theory explains why oracles and psychics are able to give good advice it begs the question of why there is any need to lie. Why do psychics lie and say they have supernatural powers? If they do provide good advice, why not just announce that to the clients? This point is better addressed with the following example.

## 1.1 Example

Consider an infinite sequence of periods, where, in each period, there is a short lived agent (he) who must make a decision $a_t \in \{L, R\}$. The goal of the agent is to match the random

[3] See Fontenrose (1978).

state of the period $\omega_t \in \{L, R\}$. If he does, he receives a payoff of 1, if he does not his payoff is 0. Before making the decision each agent observes a signal $s_t \in \left\{\overline{\alpha}, \overline{\beta}, \underline{\alpha}, \underline{\beta}\right\}$. Assume that, for all $t \geq 0$,

$$\Pr\{s_t = s\} = \frac{1}{4} \text{ for all } s \in \left\{\overline{\alpha}, \overline{\beta}, \underline{\alpha}, \underline{\beta}\right\}$$

and that

$$\Pr\{\omega_t = L | s_t = \overline{\alpha}, s_{t-1}\} = \begin{cases} 0.9 \text{ if } s_{t-1} = \overline{\alpha} \\ 0.1 \text{ if } s_{t-1} = \overline{\beta} \\ 0.7 \text{ if } s_{t-1} = \underline{\alpha} \\ 0.3 \text{ if } s_{t-1} = \underline{\beta} \end{cases}$$

while

$$\Pr\left\{\omega_t = L | s_t = \overline{\beta}, s_{t-1}\right\} = 1 - \Pr\{\omega_t = L | s_t = \overline{\alpha}, s_{t-1}\}$$

and

$$\Pr\{\omega_t = L | s_t, s_{t-1}\} = \frac{1}{2} \text{ if } s_t \in \left\{\underline{\alpha}, \underline{\beta}\right\}$$

for all $s_{t-1} \in \left\{\overline{\alpha}, \overline{\beta}, \underline{\alpha}, \underline{\beta}\right\}$ and for any $t \geq 1$. Notice that, for the agent of period $t$, observing only his signal $s_t$ does not help him because

$$\Pr\{\omega_t = L | s_t = s\} = \frac{1}{2} \text{ for all } s \in \left\{\overline{\alpha}, \overline{\beta}, \underline{\alpha}, \underline{\beta}\right\}$$

However, if $s_{t-1}$ is also known one can identify two classes of signals. If the signal is "bad", which is the case when $s_t \in \left\{\underline{\alpha}, \underline{\beta}\right\}$, nothing really changes for the agent. However, if the signal is "good", which is the case when $s_t \in \left\{\overline{\alpha}, \overline{\beta}\right\}$, the probability that the agent is able to match the state $\omega_t$ is either 0.7 or 0.9, depending on $s_{t-1}$. So, at least when $s_t \in \left\{\overline{\alpha}, \overline{\beta}\right\}$, there is an interest by agent $t$ to access the information held by agent $t-1$.

To facilitate the exchange of information, assume that there is a long lived agent called the expert (she). Assume first that the only thing that the expert does is to collect all the signals of his past clients and give as good advice as possible to any client who might consult her. The first problem she faces is the first client, because there is no advantage for him to pay $c$ to learn nothing. To bypass this problem, assume that the expert is in possession of $s_0$. Given this, consider the problem of agent $t = 1$. While he may consult the expert if $s_1 \in \left\{\overline{\alpha}, \overline{\beta}\right\}$ and $c$ is not too large, he certainly will not if $s_1 \in \left\{\underline{\alpha}, \underline{\beta}\right\}$. In fact, no agent at any period $t$ will ever consult the expert if $s_t \in \left\{\underline{\alpha}, \underline{\beta}\right\}$. Assume that each agent observes which of the preceding agents consulted the expert and, if they did, whether they were able to match the state. In this case, at the moment that some agent chooses not to consult the expert (which happens at least when $s_t \in \left\{\underline{\alpha}, \underline{\beta}\right\}$), no one else will ever consult the expert again. So, even if $c$ is arbitrarily small, the chain of clients who consult the expert will break almost surely. This is bad for the expert, who will not be able to attract as many clients as she would like, and is bad for the agents because, if $c$ is small enough, every agent would be better off (ex-ante) if everyone blindingly chose to consult the expert at every period, no matter what.

Imagine that, at period $t = 0$, the expert announces to the agents that the world is not what they think it is. In particular, the lie that the expert tells is that $\omega_t$ does not depend on $s_t$ and $s_{t-1}$ but rather only on some signal $\eta_t \in \{L, R\}$ that only she observes in every period. In particular, the expert says that $\Pr\{\eta_t = L\} = \frac{1}{2}$ for all $t$ and that

$$\Pr\{\omega_t | \eta_t\} = \begin{cases} 0.65 \text{ if } \omega_t = \eta_t \\ 0.35 \text{ if } \omega_t \neq \eta_t \end{cases}$$

for all $t \geq 1$.

Agents are not assumed to blindly believe the expert. What is assumed instead is that they place a strictly positive probability $\pi$ on the lie being true and then, based on the past success of the expert, update their beliefs. In this way, the lie that the expert tells must be empirically indistinguishable from the truth, in order to keep the lie going.

Assume that $\pi = \frac{1}{2}$ and continue to assume that $s_0$ is observed by the expert. Consider agent $t = 1$. If his signal is $s_t \in \{\underline{\alpha}, \underline{\beta}\}$, he will consult the expert if and only if

$$\frac{1}{2}\left(0.65 - \frac{1}{2}\right) - c \geq 0$$

So, if $c \leq 0.075$, agent $t = 1$ will consult the expert no matter what his signal is. This is not particularly impressive because it is not hard to come up with a lie which convinces the first agent to come in. The only thing necessary is to make the lie sufficiently appealing. The real issue is to make the lie believable and still appealing.

Consider what happens at period $t = 2$. The agent at period $t = 2$ is assumed to observe that agent $t = 1$ consulted the expert and, because he did, whether he was able to match the state. So, at period $t = 2$, he will update his belief about the lie. In order to do so, he must calculate the probability that agent $t = 1$ was successful under the two hypothesis: that the world is such that $\omega_t$ depends on $s_t$ and $s_{t-1}$ and that the world is such that $\omega_t$ depends on $\eta_t$. However, notice that, under both hypothesis, the probability of success is the same and is equal to 0.65. Therefore, at period $t = 2$, the belief will be the same as in period $t = 1$. Basically, the lie that is being told promises a success rate that is, on average, the same as in the real world, so that it becomes empirically indistinguishable from it. So, at period $t = 2$, the agent will again consult the expert, which will generate the same belief for period 3 and so on.

Simply put, the expert lies because it attracts more clients. If the expert did not lie, at period $t = 1$, the agent would sometimes be very willing to consult her (if $s_1 \in \{\overline{\alpha}, \overline{\beta}\}$) but some other times would be very unwilling (if $s_1 \in \{\underline{\alpha}, \underline{\beta}\}$). The problem for the expert is that, when the agent is very unwilling, he will not consult her, which will have multiplying effects going forward, because, as a consequence, no future agent would consult her. So, it is in the expert's interest to increase the agent's willingness to consult when he is less willing ($s_1 \in \{\underline{\alpha}, \underline{\beta}\}$). In order to do this and still have a "sustainable" lie it is necessary to reduce the agent's willingness to consult when he is more willing ($s_1 \in \{\overline{\alpha}, \overline{\beta}\}$). A lie which makes the past less relevant accomplishes exactly that.

Furthermore, the "lie" not only makes the expert better off but also the agents. In particular, if $c$ is small, the expert is able to tell a lie that gets every agent to consult her, which increases everyone's utility ex-ante (except agent $t = 1$). In fact, the main result of the paper is that, if all agents would always be better off by committing to consulting the expert before observing their own signal, it is possible to build a lie which accomplishes just that and is empirically indistinguishable from the truth.

Finally, notice that, while I have assumed that the lie is initially believed with a probability of 50%, that is not necessary. First, because, if the expert was less convincing, the consequence would be that the threshold over $c$ would be smaller but the same result would go through. And second, as I discuss in section 5, even if $c$ is larger, it is possible that the lie becomes permanent at least with a positive probability.

### 1.2 Related Literature

One of the things that I do in this paper is to show how an expert can create a lie which is empirically indistinguishable from the truth, and in that sense, lasts forever. There is a literature, initiated by Foster and Vohra (1998) among others, which discusses the somewhat related problem of whether it is possible to design a test that empirically validates someone's theory. Imagine that an expert claims to know the distribution of some random variable. An uninformed agent would like to have a way to test whether the claim of the expert is correct. Is there such a test? The challenge is that, while the expert might be lying and, may, in fact, have no idea about the distribution of the random variable, knowing the test allows her to strategically tailor her predictions in order to pass. The general message of this literature seems to be that it is virtually impossible to design such tests, at least if one considers only tests which only use past predictions of the expert.[4] While the two problems are fundamentally alike - whether it is possible that someone makes up a lie which lasts forever despite there being an unlimited amount of data available - the two approaches are different. In the above literature, the test is announced from the outset, the expert knows what it is and then makes her predictions in order to pass it, without being concerned about how her actions influence the uninformed agent's beliefs. In my paper, by contrast, the only things that matter are the beliefs of the uninformed agents because those are what determine whether they choose to consult the expert. In this sense, my approach fits more with the

[4] See Dekel and Feinberg (2006), Olszewski and Sandroni (2007) and Olszewski and Sandroni (2008).

literature on reputations in repeated games, which I discuss below. The expert in my paper creates a lie which enables her to generate a reputation that such a lie is true. Nevertheless, the message of my paper is very much aligned with this literature on the empirical testing of theories: it is at least possible for an expert to put forward a false theory about the world and never be disproved despite there being unlimited evidence.

There is an extensive literature on reputations in Economics. The general idea of this literature is that "normal" agents can create reputations of being of a "commitment" type.[5] Cripps, Mailath and Samuelson (2004) consider a model where a long-run agent faces an infinite sequence of short-lived agents in a context with imperfect monitoring. They show that, for any equilibrium, the normal long-lived agent is unable to sustain the belief that he is the commitment type, so that there are no permanent reputations. Several of the papers that followed describe conditions under which the lie that the normal type is the commitment type may last asymptotically: for example, if the type of the long-run player is not permanent (Mailath and Samuelson (2001) and Ekmeci, Gossner and Wilson (2012); or if the access to past data is either costly (Liu (2011)) or limited (Ekmekci (2011) and Hu (2016)).

In my paper, the reputation is not about the long-lived agent's type but about the world, i.e. the distribution of the public signal. The basic idea of the reputation literature is that, if it is conceivable that a commitment type exists, a normal type agent can sustain the belief that he is the commitment type through his actions and become better off as a result. The only difference in my paper is that, instead of the long-lived agent using the fact that it is conceivable that a commitment type exists, he uses the fact that it is conceivable that an alternative world exist. In a way, I replace the commitment type by a false world, and then study a similar question: can the long-lived agent do better by building up a reputation that the false world is true?

# 2 Model

## 2.1 Description of the game and the payoffs

Consider a model with infinitely many periods. In each period $t$, there is a short lived agent $t$ (he) who must take an action $a_t \in \{L, R\}$. The agent receives a payoff of 1 if his action matches the state of the world $\omega_t \in \{L, R\}$, which is random and unobservable. Otherwise, the agent receives no payoff. Let

$$u_t = \begin{cases} 1 \text{ if } a_t = \omega_t \\ 0 \text{ if } a_t \neq \omega_t \end{cases}$$

[5]See Kreps, Milgrom, Roberts and Wilson (1982), Kreps and Wilson (1982) and Milgrom and Roberts (1982).

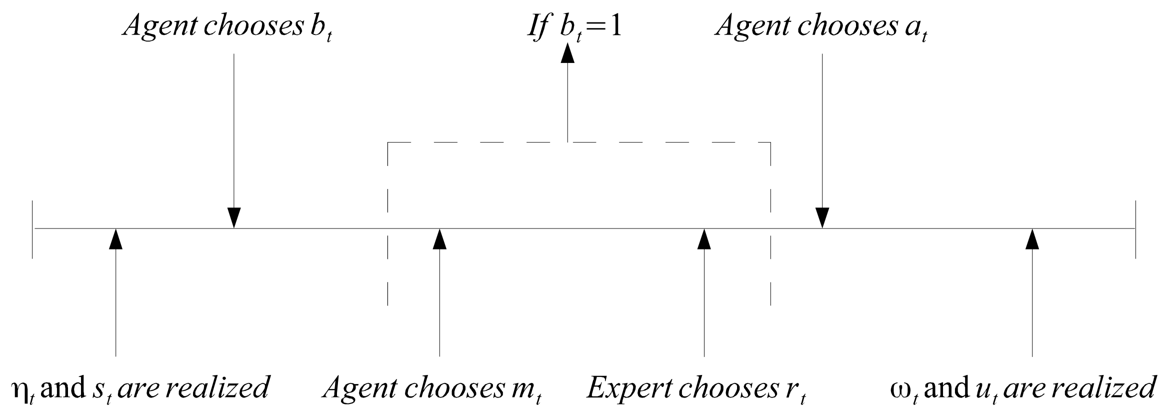


Figure 1: Timing within any period $t$

If $u_t = 1$, I refer to the agent has having been successful. Before deciding $a_t$, each agent $t$ is assumed to observe a private signal $s_t \in [0, 1]$.

There is a long-lived agent called the expert (she). The expert is assumed to observe a private signal $\eta_t \in \{L, R\}$ in every period $t$ in addition to have access to more past information than the agent (see the section on information). Each agent may choose to consult the expert before deciding $a_t$ (but after observing $s_t$) at a cost $c > 0$. Let

$$b_t = \begin{cases} 1 \text{ if agent } t \text{ consults the expert} \\ 0 \text{ if agent } t \text{ does not consult the expert} \end{cases}$$

If the agent consults the expert, he sends a message $m_t \in M_t$ to the expert, where $M_t$ is assumed to be arbitrarily large. After observing $m_t$, the expert responds by sending a recommendation $r_t \in \{L, R\}$ to the agent. The agent is then free to follow the advice of the expert or not. Figure 1 displays the timing of the events within each period.

The distribution of all the random variables $\{\omega_t, s_t, \eta_t\}_{t=1}^{\infty}$ depends on a single random variable $\theta \in \{T, F\}$, where $T$ stands for "true" and $F$ stands for "false", which is realized at the beginning of period 0 and is **only** observed by the expert. Let $\pi \in [0, 1)$ denote the probability that $\theta = F$.

The interpretation of this modelling choice is the following. At period 0, the expert observes the "true" world, i.e. he becomes aware of the distribution of $\omega_t$, $s_t$ and $\eta_t$. However, she is able to tell a "lie", i.e. she is able to claim that the world works in a different way. At period 0, the public belief is that there is a probability $\pi$ that such alternative world is real. Therefore, if $\theta = R$, the distribution of $\{\omega_t, s_t, \eta_t\}_{t=1}^{\infty}$ is the true one, whereas if $\theta = F$, it is the one that the expert says it is at the beginning of period 0.

Finally, it is assumed that the expert wants to maximize her discounted expected profit:

$$E_{\{b_t\}_{t=0}^{\infty}}\left[\sum_{t=0}^{\infty}\delta^t b_t c\right]$$

where $\delta \in (0,1)$. I am particularly interested in the case where $\delta \to 1$.

## 2.2 Information

The public history $h^t$ available at the beginning of period $t$ is

$$h^t = \{b_\tau, b_\tau u_\tau\}_{\tau=1}^{t-1}$$

and is observable by every agent $t' \geq t$. In words, each agent $t$ observes whether the agents that came before him consulted the expert and, if so, whether they were successful. In addition to the public history $h^t$, agent $t$ observes $s_t$ before deciding whether to consult the expert. If he chooses to consult the expert, he also observes her report $r_t$ before deciding $a_t$.

At the beginning of period $t$ (or at the end of period $t-1$), the expert is assumed to have observed i) which of the previous agents consulted with her, ii) their messages, iii) her own past recommendations, iv) whether the past agents who consulted with her were successful and v) all past private signals. Formally, let

$$\widehat{h}^t = \{b_\tau, b_\tau m_\tau, b_\tau r_\tau, b_\tau u_\tau, \eta_\tau\}_{\tau=1}^{t-1}$$

so that, at the beginning of period $t$, the expert observes $\widehat{h}^t$. During the period, she is assumed to observe $\eta_\tau$, which is assumed to be realized before the agent has had the chance to consult with the expert, and message $m_t$ if agent $t$ chooses to consult her. Finally, recall that the expert is always informed about $\theta$.

## 2.3 The True World

If $\theta = T$, it is assumed that, for all $t \geq 1$, each $s_t \sim U(0,1)$, while $\omega_t$ depends only on $s_t$ and $s_{t-1}$. In particular, I assume that there is a function $p : [0,1]^2 \to [0,1]$ such that

$$\Pr\{\omega_t = L | s_t, s_{t-1}\} = p(s_t, s_{t-1})$$

for all $t$, $s_t \in [0,1]$ and $s_{t-1} \in [0,1]$.

As for $\eta_t$, it is assumed to be independent of any $\omega_\tau$ and $s_\tau$ for any $\tau \geq 0$, i.e. it is completely uninformative for any agent.

In the real world, the only information an agent at period $t$ should care about is contained in his own signal and in the previous agent's signal. The benefit of the expert in this context is that she aggregates information, i.e. she makes past information available to current agents.

Notice that the assumption that $s_t \sim U(0,1)$ is without loss of generality because the function $p$ has no restrictions. In particular, it is possible to think of signals as being discrete as in the example, by simply properly designing $p$. In the case of the example, it follows that

$$p(s_t, s_{t-1}) = \begin{cases} 0.9 \text{ if } s_{t-1} \leq 0.25 \\ 0.1 \text{ if } s_{t-1} > 0.75 \\ 0.7 \text{ if } s_{t-1} \in (0.25, 0.5] \\ 0.3 \text{ if } s_{t-1} \in (0.5, 0.75] \end{cases} \quad \text{if } s_t \leq 0.25$$

and

$$p(s_t, s_{t-1}) = \begin{cases} 0.1 \text{ if } s_{t-1} \leq 0.25 \\ 0.9 \text{ if } s_{t-1} > 0.75 \\ 0.3 \text{ if } s_{t-1} \in (0.25, 0.5] \\ 0.7 \text{ if } s_{t-1} \in (0.5, 0.75] \end{cases} \quad \text{if } s_t > 0.75$$

while $p(s_t, s_{t-1}) = 0.5$ if $s_t \in (0.25, 0.75]$.

## 2.4 The False World

If $\theta = F$, it is assumed that, for all $t \geq 1$

$$\Pr\{\omega_t = L\} = \frac{1}{2} \text{ and } s_t \sim U(0,1)$$

As for the signal $\eta_t$ that the expert observes, it is assumed that, for all $t \geq 1$,

$$\Pr\left\{\eta_t = \omega_t | \omega_t, s_t, h^t\right\} = \frac{1}{2} + \lambda_t^{h^t}(s_t)$$

where $\lambda_t^{h^t}(s_t) \in \left[0, \frac{1}{2}\right]$ for all $s_t \in [0,1]$ and any history $h^t$.

The main difference from the false world to the true world is that in the false world, $s_{t-1}$ does not influence directly $\omega_t$. So, while in the real world, the agent is concerned with learning about $s_{t-1}$, in the false world, he is concerned with learning about $\eta_t$ (because he already observes $h^t$). In the false world, the expert is helpful not because he makes past information available but because he has access to a relevant source of information herself. Therefore, $\eta_t$ works as a "signal from the Gods".

This formulation for the false world allows the lie to be dependent of the public history and of the agent's signal. If $\lambda$ is independent of the history, I say that the lie is "simple". In section 4, I study these simple lies in more detail. In the example, the lie was not only independent of the history but also of the agent's signal, as the expert promised a constant success rate of 65%.

Lies that are dependent on the public history, while perhaps more farfetched, seem compatible with what one might expect to hear from any number of cults or psychics. For example, the lie could be that the signal that the expert receives comes straight from God Xenu. But God Xenu is a rather petty God who does not want the expert to do too well. So, while, normally the signal that the expert obtains from God Xenu is rather informative (say $\lambda_t^{h^t}(s_t) = 0.25$), if she has provided successful advice in the preceding 5 periods, God Xenu will only supply the expert with a bad signal in the 6th period in the form of a $\lambda_t^{h^t}(s_t) = 0$.

Finally, notice that, for all $h^t$ and $s_t \in [0, 1]$,

$$\Pr\left\{\omega_t | s_t, h^t\right\} = \Pr\left\{\eta_t | s_t, h^t\right\} = \frac{1}{2} \text{ for all } \omega_t \in \{L, R\} \text{ and } \eta_t \in \{L, R\}$$

so that

$$\Pr\left\{\omega_t = L | s_t, h^t\right\} = \frac{1}{2}$$

Therefore, an agent who is born at period $t$ and only observes the history $h^t$ and his own signal $s_t$, is indifferent between choosing $L$ and $R$ if he believes that the false world has been realized. So, the structure of the lie makes the opportunity cost of consulting the expert in the false world as small as possible. If that same agent had access to $\eta_t$, the probability of success would be $\frac{1}{2} + \lambda_t^{h^t}(s_t)$ so that $\lambda_t^{h^t}(s_t)$ represents how much the agent would be willing to pay to access the signal of the expert in the false world.

## 2.5 Period $t = 0$

Notice that, if $\theta = T$, the expert faces a problem immediately at period 0. Recall that, if every agent is aware that they are in the real world, the only reason why they would want to consult the expert would be to access past information that previous agents have provided the expert with. But at period 0, it would be known that the expert has no information from past periods, because it is assumed that there are no past periods. As a result, agent $t = 0$ would never consult the expert and, as a chain reaction, no one would ever consult the expert.

To avoid this, I assume that the expert holds $s_0$ in her possession (so that there is no agent $t = 0$), in addition to assuming that there is no $u_0$ (or that it is not observable). In this way, the agent at $t = 1$ knows only that $s_0$ is known by the expert.

### 2.6 Strategies and Equilibrium Concept

For any $t$, agent $t$'s strategy is made of three components: i) a choice of whether to consult the expert, ii) a choice of the message to relay to the expert should he consult her and iii) what action to take. Let $\widetilde{m}_t^{h^t}(s_t) \in M_t$ denote the choice of agent $t$ of what message to send to the expert after public history $h^t$ and signal $s_t$, should he choose to consult her. I say that agent $t$ reports truthfully at history $h^t$ if $\widetilde{m}_t^{h^t}(s_t) = s_t$ for any $s_t \in [0,1]$. Agent $t$ is said to follow a truthful reporting strategy if he reports truthfully after any public history.

The expert's strategy is simply a choice on what recommendation to give, for any (private) history $\widehat{h}^t$, message $m_t \in [0,1]$ sent by agent $t$, signal $\eta_t \in \{L, R\}$ and $\theta \in \{L, R\}$.

A perfect Bayesian equilibrium (PBE) is a strategy for the expert and a strategy for each agent such that i) the expert chooses her recommendation optimally given her beliefs and the agents' strategies, ii) each agent makes each decision optimally given their beliefs and the other agents and expert's strategies and iii) beliefs are updated according to Bayes' rule whenever possible.

## 3 Analysis

### 3.1 If there is no lie

Assume first that $\pi = 0$ so that the expert is honest. In this case, every agent knows that $\theta = T$: everyone knows that the world is the true world.

Let

$$v(s) = \int_0^1 \max\{p(s,x), 1 - p(s,x)\}\, dx$$

and notice that it represents the expected benefit of consulting the expert for agent 1 when $s_1 = s$ and the expert is as helpful as possible, i.e. the expert reports $L$ if and only if

$$\Pr\{\omega_1 = L | s_0, s_1, \theta = T\} \geq \frac{1}{2} \text{ for all } (s_0, s_1) \in [0,1]^2$$

Agent 1's opportunity cost of consulting the expert is given by

$$q(s) = \max\left\{\int_0^1 p(s,x)\, dx, 1 - \int_0^1 p(s,x)\, dx\right\}$$

Finally, let

$$g(s) \equiv v(s) - q(s)$$

It follows that, if agent 1 draws some signal $s_1 = s$ such that

$$g(s) < c$$

the agent will not consult the expert, no matter how the expert decides her report. Let

$$I^c = \{s \in [0,1] : g(s) < c\}$$

and notice that the probability that agent 1 does not consult the expert is at least $\int_{s \in I^c} ds$.

**Proposition 1** *If $p$ is such that*

$$\int_{s \in I^c} ds > 0 \tag{1}$$

*then*

$$b_t \to^{a.s.} 0$$

*in any PBE.*

**Proof.** See Appendix. ■

Notice that once the chain of agents who consults the expert is broken, it is never restored: if an agent does not consult the expert, there is no advantage whatsoever to consulting the expert for the following agent. Therefore, if the condition holds, and agent 1 does not consult the expert, no one else will. If $s_1$ is such that the agent does consult the expert, it is possible that agent 2 is interested in consulting the expert as well. However, on average, he is less inclined to do so than agent 1 was because the expert will have less relevant information at period 2 than at period 1 (recall that $s_0$ is assumed to be known by the expert). As a result, even if agent 1 consults the expert, there is a positive probability that agent 2 prefers not to. In proposition 1, I show that, for any PBE, the probability that the chain is broken at period $t + 1$ given that it was not broken until period $t$, for any $t$, is larger than some $\varepsilon > 0$. Therefore, the probability that agent $t$ consults the expert converges to 0 as $t$ becomes increasingly large.

## 3.2 Main Result

The problem of the expert is to find a lie, or more specifically $\lambda$, which enables her to do better than what she was able to do if she was honest. In this section, I find sufficient conditions on function $p$ for which there is such a lie which not only makes the expert better off, but actually maximizes her profit, i.e. gets every agent to consult her in every period. Moreover, such conditions are less restrictive than condition (1). In other words, as the example of section 1.1 illustrates, it is possible that condition (1) holds but, at the same time, there is a lie which makes every agent consult the expert in every period.

My approach is the following: I select $\lambda$ so that a PBE with the following properties exists:

$A$ : every agent consults the expert in every period for any signal,
$B$ : the public belief that $\theta = F$ is always $\pi$.
$C$ : the expert gives as good advice as she is able to on the path of play,
$D$ : the expert gives uninformative advice off the path of play,
$E$ : every agent reports truthfully.

The path of play of the described profile is the path of public histories for which every preceding agent has consulted the expert. Let $H^t$ be the set of all public histories on the path of play at period $t$. It follows that

$$h^t \in H^t \Leftrightarrow b_\tau | h^t = 1 \text{ for all } \tau < t$$

where $b_\tau | h^t$ denotes the choice of agent $\tau$ of whether to consult the expert consistent with history $h^t$.

Property $C$ is a consequence of property $A$ in that, if the expert gives as good advice as she is able to, she obtains the maximum possible profit so any deviation from this behavior would not be strictly beneficial to her. Formally, to "give as good advice as she is able to" on the path of play means that, for any private history $\widehat{h}^t$ consistent with any public history $h^t \in H^t$, the choice of the report of the expert is $L$ if and only if

$$\Pr\left\{\omega_t = L | \widehat{h}^t, m_t, \eta_t, \theta\right\} \geq \frac{1}{2}$$

for all $m_t \in M$, $\eta_t \in \{L, R\}$ and $\theta \in \{T, F\}$.

Property $D$ ensures that no agent consults the expert off the path of play. Basically, it says that in any private history $\widehat{h}^t$ not consistent with any $h^t \in H^t$, the expert reports $L$ with a probability of 50%, for any $m_t \in M$, $\eta_t \in \{L, R\}$ and $\theta \in \{T, F\}$. The expert does not want to deviate from this for the same reason that, in standard static cheap talk games, there is always a babbling equilibrium.[6]

Finally, agents want to report truthfully (property $E$) because it helps the expert give better advice.

The challenge then becomes one of finding such a lie for which, whenever properties $C, D$ and $E$ hold, then properties $A$ and $B$ also hold. Before stating the result, it is necessary to introduce additional notation.

[6]See Crawford and Sobel (1982).

Consider some agent $t$ at some public history $h^t \in H^t$ after observing some signal $s_t = s$. At this point, the agent will have a belief about $\theta$. If $\theta = T$, the agent's expected benefit of consulting the expert is equal to $\frac{1}{2} + \lambda_t^{h^t}(s)$. If, however, $\theta = F$, the expected benefit of the agent, which I denote by $v_t^{h^t}(s)$, will depend on what the previous agent has chosen to do. By property $A$, agent $t$ is able to infer that the previous agent would have consulted the expert for any $s_{t-1}$. However, observing whether agent $t-1$ was successful will allow agent $t$ to update his beliefs about $s_{t-1}$ so that, in general, he will not believe that $s_{t-1}$ is distributed uniformly. One can recursively define $v_t^{h^t}$ as follows:

$$v_1^{h^1}(s) = v(s) \text{ for all } s \in [0,1]$$

while, for any $h^t \in H^t$ and any $t > 1$,

$$v_t^{h^t}(s) = \int_0^1 d_t^{h^t}(x) \max\{p(s,x), 1 - p(s,x)\}\, dx$$

where

$$d_t^{h^t}(x) = \begin{cases} \dfrac{v_{t-1}^{h^{t-1}}(x)}{\int_0^1 v_{t-1}^{h^{t-1}}(x)dx} & \text{if } u_{t-1}|h^t = 1 \\ \dfrac{1-v_{t-1}^{h^{t-1}}(x)}{\int_0^1 \left(1-v_{t-1}^{h^{t-1}}(x)\right)dx} & \text{if } u_{t-1}|h^t = 0 \end{cases}$$

The density function $d_t^{h^t}$ represents the beliefs that agent $t$ has about $s_{t-1}$ when at history $h^t \in H^t$. In an analogous way, define $q_t^{h^t}(s)$ to denote the expected payoff for agent $t$ of not consulting the expert if $\theta = T$, conditional on any history $h^t \in H^t$ and on $s_t = s$, which can be written as

$$q_t^{h^t}(s) = \max\left\{\int_0^1 d_t^{h^t}(x)\, p(s,x)\, dx, \int_0^1 d_t^{h^t}(x)\,(1 - p(s,x))\, dx\right\}$$

where

$$d_1^{d_t^{h^1}}(x) = 1 \text{ for all } x \in [0,1]$$

Finally, let

$$g_t^{h^t}(s) = v_t^{h^t}(s) - q_t^{h^t}(s)$$

for all $h^t \in H^t$ and for all $t$.

**Proposition 2** *If, for all $h^t \in H^t$ and for all $t$,*

$$\int_0^1 g_t^{h^t}(s)\,ds \geq c \tag{2}$$

*and*

$$\int_0^1 v_t^{h^t}(s)\,ds \leq k < 1 \tag{3}$$

*for some $k$, there is $\lambda$ and $\pi \in (0,1)$ such that there is a PBE with properties A through E.*

**Proof.** See Appendix. ■

Condition (3) is a technical condition which simply states that the expected benefit of consulting the expert can be bounded away from 1. For it to be satisfied it is enough, for example, that $p$ is bounded away from 1.

The most meaningful condition is (2). Notice that, by definition, $g$ is completely independent of the lie that the expert tells. What the condition states is that, on the path of play, the average expected net benefit of consulting the expert, even if $\theta = T$, is larger than $c$. To better understand condition (2), consider the following thought experiment. Suppose that each agent must decide whether to consult the expert **before** observing their own signal. What condition (2) implies is that, in such a case, even if it was known that $\theta = T$, each agent would prefer to consult the expert for any history where the previous agent also did so. So, in this hypothetical scenario, there would be no need for the expert to lie. Every agent would consult the expert in every period, the expert would be maximizing her profit and all agents would be better off than if there was no expert. What the lie does is it transforms the actual problem of the expert into this hypothetical problem.

Notice also that it is possible that both conditions (1) and (2) hold. When this happens, there is a clear benefit, at least to future agents, from the expert's lie. If the expert is honest, the probability that agents born in the arbitrarily far away future will have a positive utility is 0. But, with a properly chosen lie, they will have a positive expected utility.

The idea of the argument is the following. In order for the belief about $\theta$ to be constant, it must be that the probability of success in the true and in the false worlds is the same, when consulting the expert. So, for any public history $h^t$ on the path of play, it must be that

$$\int_0^1 v_t^{h^t}(s)\,ds = \int_0^1 \left(\lambda_t^{h^t}(s) + \frac{1}{2}\right) ds \tag{4}$$

where the LHS refers to the true world and the RHS to the false world. By guaranteeing that the **average** expected benefit of consulting the expert is the same in both worlds, the true and the false, it becomes impossible for agents in future periods to distinguish the two.

Furthermore, for each agent to want to consult the expert after said public history, it must be that

$$\pi \lambda_t^{h^t}(s) + (1-\pi)\, g_t^{h^t}(s) \geq c \tag{5}$$

Because of conditions (2) and (4), condition (5) holds **on average**. But then, it is possible to choose $\lambda_t^{h^t}$ so that condition (5) holds for all $s \in [0,1]$. The logic is the same as in the example. If it is known that $\theta = T$, there will be signals for which the agent will not want to consult the expert even though, on average, he does. What the lie does is that it reduces the expected benefit of consulting the expert when the agent is very willing to consult her, in order to be able to increase the expected benefit of consulting the expert when he is not, i.e. it averages out the expected benefits.

# 4 Simple Lies

Simple lies are defined to be history independent lies. Formally, the expert tells a simple lie if $\lambda_t^{h^t}$ is independent of $h^t$ and of $t$.

From condition (5), one can see that the history dependence that a lie must have in order for proposition 2 to hold comes from the fact that $g_t^{h^t}$ might be history dependent. So, for the lie to be history independent, it is sufficient that $g_t^{h^t}$ is also history independent. When that happens, condition (2) becomes

$$\int_0^1 g(s)\, ds \geq c$$

which enables an even clearer comparison with condition (1). Recall that, from section 3.1, unless

$$\int_0^1 \mathbf{1}\{g(s) \geq c\}\, ds = 1$$

agents will almost surely stop consulting the expert. Thus, unless $g$ is constant, there is always some interval $\widehat{I} \in \mathbb{R}^+$ such that, if $c \in \widehat{I}$, the expert has a strict incentive to tell the lie of proposition 2.

**Proposition 3** *If $g_t^{h^t}$ is independent of $h^t$ and $t$, and*

$$\int_0^1 \mathbf{1}\left\{ g(s) < \int_0^1 g(x)\, dx \right\} ds > 0$$

*there is an interval $\widehat{I} \in \mathbb{R}^+$ such that, for all $c \in \widehat{I}$, conditions (1) and (2) both hold.*

**Proof.** See Appendix. ■

When is $g_t^{h^t}$ independent of the history? The reason that $g_t^{h^t}$ might depend on the history is that the beliefs of an agent regarding the previous period's signal evolve over time as more and more evidence is realized. Hence, in order to eliminate that dependence, it is sufficient that each agent does not change his beliefs about the previous signal after observing whether the previous agent was successful - formally, if $p$ is such that $d_t^{h^t} = d_1^{h^1}$ for all $h^t \in H^t$ and $t$.

**Proposition 4** *If $v$ is constant, then $d_t^{h^t}$ is independent of $h^t \in H^t$ and $t$.*

**Proof.** See Appendix. ■

If $v$ is constant, there is no information being transmitted at the end of the period, because a success is equally likely to happen for any signal. As a result, $d_t^{h^t} = d_1^{h^1}$ for any $h^t \in H^t$ and for any $t$ so that the problem of each agent is always the same on the path of play.

**Example 5**

$$p(s,x) = \begin{cases} k \text{ if } s \geq x \\ 1-k \text{ if } s < x \end{cases} \quad \text{for some } k \in \left[\frac{1}{2}, 1\right]$$

*In this case,*

$$v(s) = k \text{ for all } s \in [0,1]$$

It is also possible to find $p$ where, even though $d_t^{h^t}$ may **not** be history independent, $g_t^{h^t}$ will.

**Proposition 6** *If $p$ is such that i)*

$$p(s,x) \in \{k^s, 1-k^s\} \text{ for some } k^s \in \left[\frac{1}{2}, 1\right]$$

*ii)*

$$k^s = k^{1-s} \text{ for all } s$$

*and iii)*

$$p(s,x) + p(s,1-x) = \varsigma^s \text{ for some } \varsigma^s \in \left[\frac{1}{2}, 1\right]$$

*then $g_t^{h_t}$ is independent of $h^t \in H^t$ and $t$.*

**Proof.** See Appendix. ■

Finally, it is even possible that $p$ is such that the lie is not only independent of the history but also of the signal $s_t = s$. One example of such a function $p$ is the one in section 1.1. Another example, which satisfies the conditions of proposition 6 is the following:

**Example 7**

$$p(s,x) = \begin{cases} s \text{ if } x \geq \frac{1}{2} \\ 1-s \text{ if } x < \frac{1}{2} \end{cases}$$

*In this case,*

$$v_t^{h^t}(s) = v(s) = \max\{s, 1-s\}$$

*and*

$$q_t^{h^t}(s) = q(s) = \frac{1}{2}$$

*Notice that*

$$\int_0^1 v_t^{h^t}(s)\,ds = \frac{3}{4} \text{ and } \int_0^1 g_t^{h^t}(s)\,ds = \frac{1}{4}$$

*for all* $h^t \in H^t$ *and* $t$. *Notice that* $I^c > 0$ *for any* $c > 0$, *so that if* $0 < c < \frac{1}{4}$ *and the expert is honest,* $b_t \to^{a.s} 0$.

*Let*

$$\lambda_t^{h^t}(s) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

*for all* $s \in [0,1]$, $h^t \in H^t$ *and* $t$. *Notice that*

$$\pi\frac{1}{4} + (1-\pi)\left(\max\{s, 1-s\} - \frac{1}{2}\right) \geq \pi\frac{1}{4}$$

*so that, for any* $c < \frac{1}{4}$, *there is* $\pi \in (0,1)$ *for which the constant lie* $\lambda_t^{h^t}(s) = \frac{1}{4}$ *allows the expert to attract every agent in every period.*

# 5 Small Prior

In the main result, I show that, under some conditions, there is always some prior $\pi \in (0,1)$ for which there is a lie which is successful in getting every agent to consult the expert and lasts forever. Naturally, one concern is that such a $\pi$ is required to be too large. While it seems natural to me to expect that the lie is initially believed with some positive probability, it might be too optimistic to think that it will be believed with a probability close to 1. What would be the implications of restricting $\pi$ to be small?

The first observation is that it is possible to find examples where, even if $\pi$ is restricted to be small, the lie of the main result still works: it still is successful in getting every agent to consult the expert. Consider the example of section 1.1. Notice that, for some $\pi \in (0,1)$, each agent consults the expert in every period if

$$\pi\,(0.65 - 0.5) \geq c \Leftrightarrow \frac{\pi}{c} \geq 0.15$$

Therefore, for this example, what matters in terms of making the lie successful is the ratio $\frac{\pi}{c}$. Thus, it follows that, for every $\pi \in (0,1)$, there is an interval of costs for which the expert can do better by lying (recall that, if the expert is honest, agents eventually stop consulting her almost surely).

Fix some $p$ and $c$ such that conditions (1) and (2) hold but assume that $\pi < \overline{\pi}$, which is defined to be the required lower bound on the prior described in Proposition 2. While it may not be possible, in such circumstances, for the expert to succeed in getting every agent to consult her with certainty, it might be with some positive probability as I illustrate with the example of section 1.1.

Suppose that $c = 0.075$ so that, for the lie described in section 1.1 to get every agent to consult the expert in every period, it must be that

$$\pi \geq \overline{\pi} \equiv \frac{0.075}{0.15} = 0.5$$

Assume instead that $\pi = 0.05$. The problem that the expert faces is that the prior belief is too low. This does not imply that the expert should settle to being honest, because that will imply a long term average profit of 0 almost surely (if $\delta \rightarrow 1$). Instead, it is in the best interest of the expert to "build up" the belief that $\theta = F$ until it reaches $\overline{\pi}$ and then revert to the original lie - that $\lambda_t^{h^t}(s) = 0.15$ for all $s, h^t$ and $t$.

Consider the following lie:

1) For any period $t \leq 16$, choose $\lambda_t^{h^t}(s) = \frac{1}{2}$ for any $s \in [0,1]$ and for any public history $h^t$.

2) For any period $t > 16$, choose $\lambda_t^{h^t}(s) = 0.15$ for any $s \in [0,1]$ and for any public history $h^t$.

Consider a PBE where, for every history such that every preceding agent has consulted the expert and has been successful, the expert gives as good advice as she is able to, while she reports uninformatively otherwise. As for the agents, they report truthfully. This means that, once some agent refuses to consult the expert, or receives unsuccessful advice from her, no agent will ever consult the expert again.

Consider what happens at period 1.

If $s_1 \in \{\underline{\alpha}, \underline{\beta}\}$, agent 1 does not consult the expert because

$$0.05 * \frac{1}{2} < 0.075$$

but if $s_1 \in \{\overline{\alpha}, \overline{\beta}\}$ he does because

$$0.05 * \frac{1}{2} + 0.95 * 0.3 > 0.075$$

So, the lie works half the time. Now, consider what happens at period 2, after agent 1 has chosen to consult the expert.

The decision of agent 2 will depend on his belief about $\theta$, which, in turn, will depend on whether agent 1 was successful. If agent 1 was unsuccessful, the belief that $\theta = F$ will drop to 0, because in the false world, the agent should always be successful. If, however, agent 1 was successful, then the belief that $\theta = F$ will be given by

$$\frac{0.05}{0.05 + 0.95 * 0.8} \simeq 0.0617$$

So, the belief increases from period 1 to period 2. By doing the same in the following 16 periods, the expert guarantees that each agent consults the expert at least if the realized signal is $\overline{\alpha}$ or $\overline{\beta}$ and provided that every preceding agent has consulted the expert and has been successful.

At the beginning of period 17, the public belief that $\theta = F$, given that every preceding agent has consulted the expert and has been successful will be approximately 0.6.[7] Therefore, at period 17, the lie can revert back to being the "sustainable" one. Notice that agent 17 consults the expert even if $s_{17} \in \{\underline{\alpha}, \underline{\beta}\}$ because $0.6 > \overline{\pi}$. Provided the expert reaches this point in history, he succeeds in getting every future agent to consult her and is able to sustain the belief that $\theta = F$ forever. If $\delta \to 1$, the expected profit of the expert is bounded away from 0, which makes this choice of a lie preferable to honesty.

Notice that, in this example, the expert did not need to switch lies at period 17, i.e. she could have increased the belief even further and make it as close to 1 as she wanted. So, it is perfectly possible to have a sustainable very large belief in the supernatural, even though the odds of that happening are small. In Appendix $B$, I discuss how to generalize this algorithm of creating a lie even with a small prior, which works in a more general setting.

[7] See appendix $B$ for details.

# 6 Conclusion

The faith in the supernatural is often times attributed to ignorance. Those who believe in psychics or witches must have some behavioral bias which prevents them from realizing that they are being lied to. In this paper, I challenge this view and provide a theory that explains why rational agents may believe in the supernatural. Moreover, I also dispute the notion that psychics are necessarily bad. In fact, the lie about having supernatural powers that experts tell in my model only "works" if it helps their clients for, otherwise, rational agents would have no interest in consulting them. In a way, agents are being lied to for the greater good.

# 7 Appendix

## 7.1 Proof of Proposition 1

First, notice that if $c > \frac{1}{2}$, no agent will consult the expert. So, consider the case where $c \leq \frac{1}{2}$. Notice that if

$$\int_0^1 \mathbf{1}\left\{s \in I^c\right\} ds > 0$$

then there must be some $\kappa \in (0, c)$ such that

$$\int_0^1 \mathbf{1}\left\{s \in I^\kappa\right\} ds > 0$$

I show that there is a probability of at least

$$\varepsilon = \left(1 - \frac{\kappa}{c}\right) \int_0^1 \mathbf{1}\left\{s \in I^\kappa\right\} ds > 0$$

that, for any PBE and for any public history $h^t$, if $h^t$ is such that $b_\tau = 1$ for all $\tau < t$, then the probability that agent $t+1$ does not consult the expert is at least $\varepsilon$.

Take any PBE. I show the statement by induction.

Consider agent $t = 1$. The probability that he consults the expert is maximal when the expert truthfully reports at period $t = 1$. In that case, and by definition of $I^c$, there is a probability of

$$\int_0^1 \mathbf{1}\left\{s \in I^c\right\} ds > \varepsilon$$

of him not consulting the expert.

Now consider any agent $t$ after some public history $h^t$ such that $b_\tau = 1$ for all $\tau < t$. Notice that

$$\begin{aligned} v\left(s_{t+1}\right) &= \int_0^1 f\left(s_t|h^t\right) \max\left\{p\left(s_{t+1}, s_t\right), 1 - p\left(s_{t+1}, s_t\right)\right\} ds_t \\ &= \int_0^1 \sum_{h^{t+1}} \Pr\left\{h^{t+1}|h^t\right\} f\left(s_t|h^{t+1}\right) \max\left\{p\left(s_{t+1}, s_t\right), 1 - p\left(s_{t+1}, s_t\right)\right\} ds_t \\ &= \sum_{h^{t+1}} \Pr\left\{h^{t+1}|h^t\right\} \int_0^1 f\left(s_t|h^{t+1}\right) \max\left\{p\left(s_{t+1}, s_t\right), 1 - p\left(s_{t+1}, s_t\right)\right\} ds_t \end{aligned}$$

where $h^{t+1}|h^t$ is understood as representing a public history at period $t+1$ consistent with public history $h^t$. Notice that, for any $h^{t+1}$,

$$\int_0^1 f\left(s_t|h^{t+1}\right) \max\left\{p\left(s_{t+1}, s_t\right), 1 - p\left(s_{t+1}, s_t\right)\right\} ds_t \geq \widehat{v}^{h^{t+1}}\left(s_{t+1}\right)$$

where $\widehat{v}^{h^{t+1}}\left(s_{t+1}\right)$ denotes the expected utility of consulting the expert at public history $h^{t+1}$, given signal $s_{t+1}$ and the expert's strategy. The inequality follows because the expected utility of consulting the expert is larger if the expert gives as good advice as possible. This means that

$$v\left(s\right) \geq \sum_{h^{t+1}} \Pr\left\{h^{t+1}|h^t\right\} \widehat{v}^{h^{t+1}}\left(s\right)$$

It is also the case that

$$\begin{aligned} q\left(s_{t+1}\right) &= \max\left\{\int_0^1 f\left(s_t|h^t\right) p\left(s_{t+1}, s_t\right) ds_t, \int_0^1 f\left(s_t|h^t\right) \left(1 - p\left(s_{t+1}, s_t\right)\right) ds_t\right\} \\ &= \max\left\{\begin{array}{c} \int_0^1 \sum_{h^{t+1}} \Pr\left\{h^{t+1}|h^t\right\} f\left(s_t|h^{t+1}\right) p\left(s_{t+1}, s_t\right) ds_t, \\ \int_0^1 \sum_{h^{t+1}} \Pr\left\{h^{t+1}|h^t\right\} f\left(s_t|h^{t+1}\right) \left(1 - p\left(s_{t+1}, s_t\right)\right) ds_t \end{array}\right\} \\ &\leq \sum_{h^{t+1}} \Pr\left\{h^{t+1}|h^t\right\} \max\left\{\int_0^1 f\left(s_t|h^{t+1}\right) p\left(s_{t+1}, s_t\right) ds_t, \int_0^1 f\left(s_t|h^{t+1}\right) \left(1 - p\left(s_{t+1}, s_t\right)\right) ds_t\right\} \end{aligned}$$

Notice that, for all $h^{t+1}$,

$$\max\left\{\int_0^1 f\left(s_t|h^{t+1}\right)p\left(s_{t+1},s_t\right)ds_t,\int_0^1 f\left(s_t|h^{t+1}\right)\left(1-p\left(s_{t+1},s_t\right)\right)ds_t\right\}=\widehat{q}^{h^{t+1}}\left(s_{t+1}\right)$$

where $\widehat{q}^{h^{t+1}}\left(s_{t+1}\right)$ denotes the expected opportunity cost of consulting the expert at public history $h^{t+1}$ and given signal $s_{t+1}$. This means that

$$q\left(s\right)\leq\sum_{h^{t+1}}\Pr\left\{h^{t+1}|h^t\right\}\widehat{q}^{h^{t+1}}\left(s\right)$$

Therefore, it follows that

$$g\left(s\right)\geq\sum_{h^{t+1}}\Pr\left\{h^{t+1}|h^t\right\}\widehat{g}^{h^{t+1}}\left(s\right)$$

where

$$\widehat{g}^{h^{t+1}}\left(s\right)=\widehat{v}^{h^{t+1}}\left(s\right)-\widehat{q}^{h^{t+1}}\left(s\right)$$

Take any $s\in I^{\kappa}$. It follows that

$$c>g\left(s\right)\geq\sum_{h^{t+1}}\Pr\left\{h^{t+1}|h^t\right\}\widehat{g}^{h^{t+1}}\left(s\right)$$

Let

$$Z^{h^t}\left(s\right)=\left\{h^{t+1}|h^t:\widehat{g}^{h^{t+1}}\left(s\right)<c\right\}$$

so that one can write

$$\sum_{h^{t+1}\in Z^{h^t}(s)}\Pr\left\{h^{t+1}|h^t\right\}\widehat{g}^{h^{t+1}}\left(s\right)+\sum_{h^{t+1}\notin Z^{h^t}(s)}\Pr\left\{h^{t+1}|h^t\right\}\widehat{g}^{h^{t+1}}\left(s\right)\leq g\left(s\right)<c$$

Notice that

$$\Pr\left\{h^{t+1}\in Z^{h^t}\left(s\right)|h^t\right\}\underline{g}^{h^t}\left(s\right)+\left(1-\Pr\left\{h^{t+1}\in Z^{h^t}\left(s\right)|h^t\right\}\right)\overline{g}^{h^t}\left(s\right)\leq g\left(s\right)<c$$

where

$$\underline{g}^{h^t}\left(s\right)=\sum_{h^{t+1}\in Z^{h^t}(s)}\frac{\Pr\left\{h^{t+1}|h^t\right\}}{\sum_{h^{t+1}\in Z^{h^t}(s)}\Pr\left\{h^{t+1}|h^t\right\}}\widehat{g}^{h^{t+1}}\left(s\right)$$

and

$$\overline{g}^{h^t}\left(s\right)=\sum_{h^{t+1}\notin Z^{h^t}(s)}\frac{\Pr\left\{h^{t+1}|h^t\right\}}{\sum_{h^{t+1}\notin Z^{h^t}(s)}\Pr\left\{h^{t+1}|h^t\right\}}\widehat{g}^{h^{t+1}}\left(s\right)$$

Because

$$\underline{g}^{h^t}\left(s\right)<c\leq\overline{g}^{h^t}\left(s\right)$$

it follows that

$$\Pr\left\{h^{t+1}\in Z^{h^t}(s)\,|h^t\right\}\geq\frac{\overline{g}^{h^t}(s)-g(s)}{\overline{g}^{h^t}(s)-\underline{g}^{h^t}(s)}\geq\frac{\overline{g}^{h^t}(s)-g(s)}{\overline{g}^{h^t}(s)}=1-\frac{g(s)}{\overline{g}^{h^t}(s)}\geq1-\frac{g(s)}{c}\geq1-\frac{\kappa}{c}$$

Therefore, the probability that agent $t+1$ does not consult the psychic is at least

$$\int_0^1\Pr\left\{h^{t+1}\in Z^{h^t}(s)\,|h^t\right\}\mathbf{1}\{s\in I^\kappa\}\,ds\geq\left(1-\frac{\kappa}{c}\right)\int_0^1\mathbf{1}\{s\in I^\kappa\}\,ds=\varepsilon$$

## 7.2 Proof of Proposition 2

I first show the following lemma:

**Lemma 2.1.** *For every $\pi\in(0,1)$, if there is $\lambda_t^{h^t}(s)\in\left[0,\frac{1}{2}\right]$ such that*

$$\pi\lambda_t^{h^t}(s)+(1-\pi)\,g_t^{h^t}(s)=\pi\int_0^1\left(v_t^{h^t}(s)-\frac{1}{2}\right)ds+(1-\pi)\int_0^1 g_t^{h^t}(s)\,ds \tag{6}$$

*for all $s$, $h^t\in H^t$ and $t$, and (2) holds, then there is a PBE with properties $A$ to $E$.*

**Proof.** For property $B$ to hold, it must be that

$$\Pr\left\{\theta=F|h^t\right\}=\Pr\left\{\theta=F|h^{t+1}\right\} \tag{7}$$

for any $h^t$ and any $h^{t+1}|h^t$. If $h^t\notin H^t$, the equality follows by property $D$, so that no additional inferences are made once the chain breaks. If $h^t\in H^t$, and given property $A$, for the equality to hold it must be that

$$\int_0^1 v_t^{h^t}(s)\,ds=\int_0^1\left(\lambda_t^{h^t}(s)+\frac{1}{2}\right)ds$$

for all $h^t\in H^t$.

For property $A$ to hold, it is necessary that, for any $h^t\in H^t$ and for any $s_t=s$,

$$\pi\lambda_t^{h^t}(s)+(1-\pi)\,g_t^{h^t}(s)\geq c \tag{8}$$

Notice that

$$\int_0^1\lambda_t^{h^t}(s)\,ds=\int_0^1\left(v_t^{h^t}(s)-\frac{1}{2}\right)ds\geq\int_0^1\left(v_t^{h^t}(s)-q_t^{h^t}(s)\right)ds=\int_0^1 g_t^{h^t}(s)\,ds$$

so that integrating (8) implies that

$$\pi \int_0^1 \lambda_t^{h^t}(s)\, ds + (1-\pi) \int_0^1 g_t^{h^t}(s)\, ds \geq c$$

which means that (8) holds on average. Therefore, by making $\lambda_t^{h^t}(s)$ to be such that (6) holds, conditions (8) and (7) are satisfied. ■

Solving for $\lambda_t^{h^t}(s)$ using equation (6) implies that

$$\lambda_t^{h^t}(s) = \int_0^1 \left( v_t^{h^t}(s) - \frac{1}{2} \right) ds + \frac{(1-\pi)}{\pi} \left( \int_0^1 g_t^{h^t}(s)\, ds - g_t^{h^t}(s) \right) \tag{9}$$

What is left to show is that, if condition (3) holds, there is always some $\pi \in (0,1)$ for which there is such a $\lambda_t^{h^t}(s)$ for any $h^t \in H^t$.

Notice that

$$0 < c \leq \int_0^1 \left( v_t^{h^t}(s) - \frac{1}{2} \right) ds \leq k < 1$$

and

$$\left( \int_0^1 g_t^{h^t}(s)\, ds - g_t^{h^t}(s) \right) \in \left[ -\frac{1}{2}, \frac{1}{2} \right]$$

Therefore, if

$$c - \frac{(1-\pi)}{\pi} \frac{1}{2} \geq 0 \Leftrightarrow \pi \geq \frac{1}{2c+1}$$

then $\lambda_t^{h^t}(s)$ is guaranteed to be positive while if

$$k + \frac{(1-\pi)}{\pi} \frac{1}{2} \leq \frac{1}{2} \Leftrightarrow \pi \geq \frac{1}{2(1-k)}$$

then $\lambda_t^{h^t}(s)$ is guaranteed to be smaller than $\frac{1}{2}$. Hence, if

$$\pi > \max \left\{ \frac{1}{2c+1}, \frac{1}{2(1-k)} \right\}$$

then $\lambda_t^{h^t}(s)$ as defined in (9) is always between 0 and $\frac{1}{2}$.

### 7.3 Proof of Proposition 3

Let

$$\underline{c} = \sup \left\{ c \geq 0 : \int_0^1 \mathbf{1} \{g(s) \leq c\} \, ds = 0 \right\}$$

so that, for every $c > \underline{c}$, condition (1) holds. Let

$$\overline{c} = \int_0^1 g(s) \, ds$$

so that, for every $c \leq \overline{c}$, condition (2) holds. Notice that,

$$\int_0^1 \mathbf{1} \{g(s) < \overline{c}\} \, ds > 0$$

Therefore, there is $c'$ such that $\underline{c} < c' < \overline{c}$ such that, for all $c \in [c', \overline{c}] \equiv \widehat{I}$ such that

$$\int_0^1 \mathbf{1} \{g(s) < c\} \, ds > 0$$

### 7.4 Proof of Proposition 4

Let $k = v(s) = v_1^{h^1}(s)$ for all $s \in [0,1]$. Notice that, by definition, $d_1^{h^1}(x) = 1$ for any $x \in [0,1]$. Take any arbitrary period $t$ and any arbitrary history $h^t \in H^t$. Assume that $v_t^{h^t}(s) = k$ for all $s \in [0,1]$. It follows that

$$d_{t+1}^{h^{t+1}}(x) = \begin{cases} \frac{k}{\int_0^1 k da} = 1 \text{ if } u_t | h^{t+1} = 1 \\ \frac{1-k}{1-\int_0^1 k da} = 1 \text{ if } u_t | h^{t+1} = 0 \end{cases}$$

and so $d_{t+1}^{h^{t+1}}(x) = 1$ for any $x \in [0,1]$ and any $h^{t+1} \in H^{t+1}$ which follows $h^t$.

### 7.5 Proof of Proposition 6

Notice that it is enough to show that $g_2^{h^2} = g$ when $h^2 = \{b_1 = 1, b_1 u_1 = 1\} \equiv \overline{h}^2$, because this would imply that $g_2^{h^2} = g$ for $h^2 = \{b_1 = 1, b_1 u_1 = 0\}$ and, recursively, that $g_t^{h^t} = g$ for any $h^t \in H^t$ and $t$.

Notice that, by i),

$$v(s) = \int_0^1 \max\{p(s,x), 1 - p(s,x)\} dx = k^s$$

Let

$$\epsilon = \int_0^1 k^s ds$$

It follows that

$$v_2^{\overline{h}^2} = \int_0^1 \frac{k^x}{\epsilon} \max\{p(s,x), 1 - p(s,x)\} dx = k^s$$

Notice also that, by iii)

$$\int_0^1 p(s,x) dx = \frac{\varsigma^s}{2} \text{ for all } s \in [0,1]$$

so that

$$q(s) = \max\left\{\frac{\varsigma^s}{2}, 1 - \frac{\varsigma^s}{2}\right\}$$

It is also the case that

$$\int_0^1 \frac{k^x}{\epsilon} p(s,x) dx = \int_0^{\frac{1}{2}} \frac{k^x}{\epsilon} (p(s,x) + p(s,1-x)) dx = \int_0^{\frac{1}{2}} \frac{k^x}{\epsilon} \varsigma^s dx = \frac{\varsigma^s}{2}$$

where the first equality follows from ii), and the second equality follows from iii). Therefore,

$$q_2^{\overline{h}^2}(s) = \max\left\{\frac{\varsigma^s}{2}, 1 - \frac{\varsigma^s}{2}\right\} = q(s)$$

### 7.6 Appendix B

In this appendix, I complement the discussion of section 5. In part i), I discuss the example of section 1.1, while in part ii) I discuss how to generalize the algorithm described in section 5.

### 7.6.1 Part i)

Notice that, at period 2, the agent does not consult the expert if $s_2 \in \left\{\underline{\alpha}, \underline{\beta}\right\}$ because

$$0.0617 * \frac{1}{2} < 0.075$$

but does if $s_2 \in \left\{\overline{\alpha}, \overline{\beta}\right\}$ because

$$0.0617 * \frac{1}{2} + (1 - 0.0617) * 0.4 > 0.075$$

Notice that, in the true world, the net benefit for agent 2 of consulting the expert when $s_2 \in \left\{\overline{\alpha}, \overline{\beta}\right\}$ is 0.4 and not 0.3 because he knows that $s_1 \in \left\{\overline{\alpha}, \overline{\beta}\right\}$.

At period 3 and subsequent periods, the problem is the same until the belief that $\theta = F$ is equal to some $\widehat{\pi}$ such that

$$\widehat{\pi}\frac{1}{2} \geq 0.075 \Leftrightarrow \widehat{\pi} \geq 0.15$$

which happens at the beginning of period 12. At period 12, the agent will have a belief that $\theta = F$ of approximately 0.1587 and will choose to consult the expert for any signal. Therefore, the belief that $\theta = F$ at the beginning of period 13, conditional on every previous agent consulting the expert and being successful is

$$\frac{0.1587}{0.1587 + (1 - 0.1587) * 0.7} \simeq 0.2123$$

Once again, agent 13 consults the expert for any signal, so that the belief that $\theta = F$ at the beginning of period 14, conditional on every previous agent consulting the expert and being successful is

$$\frac{0.2123}{0.2123 + (1 - 0.2123) * 0.65} \simeq 0.293\,11$$

From this point onwards, the beliefs are updated in the same way until the beginning of period 17 where the belief is approximately 0.6.

### 7.6.2 Part ii)

In this part, for simplicity, I assume that $p(s, x) < k < 1$ for some $k > 0$ for any $(s, x) \in [0, 1]$. Fix any $p$ such that (2) holds. I generalize the procedure of section 5 to show how to create a permanent lie that has a positive probability of attracting every agent to consult the expert.

The procedure that I use in the example must be slightly altered to be able to hold more generally. After bringing the belief up, the idea of the argument is to use Proposition 2 to argue that, going forward, it is enough to switch to the "sustainable" lie. However, this requires that the belief that the first agent who is confronted with the sustainable lie has with respect to the previous signal, when $\theta = T$, is uniform, in order to make him equivalent to agent 1 in Proposition 2. In the example, this was the case. However, in general it may not be. Therefore, there needs to be a period in between the two sets of periods of the example to "reset" the belief about the previous signal.

Let $\varphi_\tau\left(h^t\right)$ denote the history of the last $\tau$ periods of public history $h^t$ so that

$$\varphi_\tau\left(h^t\right) = \{b_{t-\tau}, b_{t-\tau}u_{t-\tau}, ..., b_{t-1}, b_{t-1}u_{t-1}\}$$

Consider the following lie:

1) For any period $t < \bar{t}$, choose $\lambda_t^{h^t}(s) = \frac{1}{2}$ for any $s \in [0,1]$ and for any public history $h^t$.
2) If $t = \bar{t}$, choose $\lambda_t^{h^t}(s) = k - \frac{1}{2} + \epsilon$ for some $\epsilon \in (0, 1-k)$.
3) For any period $t > \bar{t}$, choose

$$\lambda_t^{h^t}(s) = \int_0^1 \left(v_{t-\bar{t}}^{\varphi_{t-\bar{t}}(h^t)}(s) - \frac{1}{2}\right) ds + \frac{(1-\pi)}{\pi}\left(\int_0^1 g_{t-\bar{t}}^{\varphi_{t-\bar{t}}(h^t)}(s)\, ds - g_t^{\varphi_{t-\bar{t}}(h^t)}(s)\right)$$

for any $s \in [0,1]$ and for any public history $h^t$.

As for the strategies, the PBE is such that each agent reports truthfully, while the expert reports as follows:

i) For every $t < \bar{t}$,

a) the expert gives as good advice as possible whenever all preceding agents have consulted the expert and have succeeded,

b) the expert gives uninformative advice otherwise.

ii) If $t = \bar{t}$,

a) the expert reports as good advice as possible if $\theta = F$ and all preceding agents have consulted the expert and have succeeded,

b) the expert gives uninformative advice otherwise.

iii) If $t > \bar{t}$,

a) the expert reports as good advice as possible if all agents **before** (but not including) $\bar{t}$ have consulted the expert and have succeeded,

b) the expert gives uninformative advice otherwise.

Part i) is as in the example and simply builds up the belief in the event that every preceding agent consults the expert and is successful. Let $\pi_{\bar{t}}$ be the belief at the beginning of period $\bar{t}$, when every agent has consulted the expert and was successful. Notice that

$$\pi_{\bar{t}} \geq \widehat{\pi}_{\bar{t}}$$

where $\widehat{\pi}_t$ is defined recursively as

$$\widehat{\pi}_1 = \pi \in (0,1)$$

and

$$\widehat{\pi}_{t+1} = \frac{\widehat{\pi}_t}{\widehat{\pi}_t + (1-\widehat{\pi}_t)\,k}$$

for all $t \geq 1$. This is because $p\,(s,x) < k$ for all $(s,x) \in [0,1]^2$. Notice that $\widehat{\pi}_{\bar{t}}$ is strictly increasing and $\widehat{\pi}_{\bar{t}} \to 1$ as $\bar{t} \to \infty$, so that the belief $\pi_{\bar{t}}$ can be made arbitrarily large, provided that the history where all preceding agents consult the expert and succeed has positive probability. Assume that is the case. Choose $\bar{t}$ to be large enough so that

a)

$$\widehat{\pi}_{\bar{t}}\left(k - \frac{1}{2} + \epsilon\right) \geq c \Leftrightarrow \widehat{\pi}_{\bar{t}} \geq \frac{c}{k - \frac{1}{2} + \epsilon}$$

so that, at period $\bar{t}$, agent $\bar{t}$ wants to consult the expert for any signal, provided every preceding agent has consulted the expert and succeeded.

b)

$$\widehat{\pi}_{\bar{t}} \geq \pi^*$$

where $\pi^*$ uniquely solves

$$\frac{\pi^*\,(1-k-\epsilon)}{\pi^*\,(1-k-\epsilon) + (1-\pi^*)\,\frac{1}{2}} = \max\left\{\frac{1}{2c+1}, \frac{1}{2\,(1-k)}\right\}$$

The right hand side comes from the proof of Proposition 2. It is such that if $\pi$ is larger than the RHS, the lie of Proposition 2 is successful in getting every agent to consult the expert. So, this condition guarantees that, at period $\bar{t}+1$, if all agents before $\bar{t}$ consulted the expert and were successful, the posterior belief that $\theta = F$ is still large enough for the sustainable lie to work, even if agent $\bar{t}$ is unsuccessful.

The argument is the following. First, the expert tells the appealing lie to get the belief $\pi_{\bar{t}}$ arbitrarily close to 1. Then, if it happens that every agent consults the expert and is successful, at period $\bar{t}$, the expert only gives good advice if $\theta = F$. Given that $\pi_{\bar{t}}$ is sufficiently large, agent $\bar{t}$ consults the expert for any signal. At period $\bar{t}+1$, the agent updates the beliefs about $\theta$ given the success of agent $\bar{t}$. If agent $\bar{t}$ is successful the belief will increase, if not it will decrease. However, because $\pi_{\bar{t}}$ was so large, the posterior belief will still be large enough at period $\bar{t}+1$ that the sustainable lie gets agent $\bar{t}+1$ to consult the expert. Furthermore, the beliefs of agent $\bar{t}+1$ will be that signal $s_{\bar{t}}$ is uniform if $\theta = T$. So then, by Proposition 2, every agent from then on will consult the expert and the belief will remain the same forever.

Notice that the reporting strategies of the agents and the expert form a PBE. In particular, at period $\bar{t}$, if all preceding agents have consulted the expert and have been successful, the expert is indifferent on what to report because, no matter what, every future agent will consult the expert. The only condition for this procedure to work is that there is a positive probability that the history for which every agent consults the expert and is successful until period $\bar{t}$ occurs. In that case, telling this lie gives the expert a positive probability bounded away from 0 when $\delta \to 1$ and is, therefore, preferable to being honest.

## References


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

[2] Cripps, M. W., Mailath, G. J., & Samuelson, L. (2004). Imperfect monitoring and impermanent reputations. Econometrica, 72(2), 407-432.

[3] Dekel, E., & Feinberg, Y. (2006). Non-Bayesian testing of a stochastic prediction. The Review of Economic Studies, 73(4), 893-906.

[4] Ekmekci, M. (2011). Sustainable reputations with rating systems. Journal of Economic Theory, 146(2), 479-503.

[5] Ekmekci, M., Gossner, O., & Wilson, A. (2012). Impermanent types and permanent reputations. Journal of Economic Theory, 147(1), 162-178.

[6] Fontenrose, J. E. (1978). The Delphic oracle, its responses and operations, with a catalogue of responses. Univ of California Press.

[7] Foster, D. P., & Vohra, R. V. (1998). Asymptotic calibration. Biometrika, 85(2), 379-390.

[8] Hu, J. (2016). Biased Learning and permanent reputation. Mimeo.

[9] Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. Journal of Economic Theory, 27(2), 245-252.

[10] Kreps, D. M., & Wilson, R. (1982). Reputation and imperfect information. Journal of economic theory, 27(2), 253-279.

[11] Liu, Q. (2011). Information acquisition and reputation dynamics. The Review of Economic Studies, 78(4), 1400-1425.

[12] Mailath, G. J., & Samuelson, L. (2001). Who wants a good reputation?. The Review of Economic Studies, 68(2), 415-441.

[13] Milgrom, P., & Roberts, J. (1982). Predation, reputation, and entry deterrence. Journal of economic theory, 27(2), 280-312.

[14] Moore, W. D. (2005). Three in Four Americans Believe in Paranormal. Retrieved from http://www.gallup.com/poll/16915/three-four-americans-believe-paranormal.aspx.

[15] Olszewski, W., & Sandroni, A. (2008). Manipulability of Future-Independent Tests. Econometrica, 76(6), 1437-1466.

[16] Olszewski, W., & Sandroni, A. (2009). A nonmanipulable test. The Annals of Statistics, 1013-1039.