No. 1
SEPTIEMBRE de 2016
quantil | Monografías
Edición electrónica

# Screening Multiple Potentially False Experts

Francisco Barreras
Álvaro J. Riascos

quantil | matemáticas aplicadas

Serie de Monografías Quantil, 2016-1
Edición electrónica.

SEPTIEMBRE de 2016




Carrera 7 # 77 - 07. Oficina 901, Bogotá, D. C., Colombia
Teléfonos: 3718132 – (310)6791459 – (320)8461236
E-mail: info@quantil.com.co
http://www.quantil.com.co

Impreso en Colombia – Printed in Colombia

# Screening Multiple Potentially False Experts

Francisco Barreras*
Research Department
Quantil

Álvaro J. Riascos†
Faculty of Economics
Universidad de los Andes

September 10, 2016

**Abstract**

A decision maker is presented with a theory from a self proclaimed expert about the probability of occurrence of certain events. The decision maker faces the possibility that the expert is completely ignorant about the data generating process and so she's interested in mechanisms that allow her to screen informed experts from uninformed ones.

The decision maker needs to control for type I error, however, since she's also uncertain about the true stochastic process, this gives room for uninformed experts to make strategic forecasts and ignorantly pass tests and profit from contracts. We present an original multiple expert model where a contract achieves screening of informed and uninformed experts by means of pitting experts' predictions against each other. Additionally, we present a theoretical review of the main findings in two branches of literature that attempt to solve the expert screening problem. Namely models about testing experts and models in contract theory that pursue screening of experts.

**Keywords**: Testing of multiple experts, manipulation, adverse selection

*E-mail: francisco.barreras@quantil.com.co
†E-mail: ariascos@uniandes.edu.co. We wish to thank Alvaro Sandroni and Paula Jaramillo for useful comments and conversations.

# Contents

# 1 Introduction

Consider the following problem: A decision maker receives probabilistic forecasts of a sequence of events from a self proclaimed expert. The decision maker would like a mechanism to evaluate such forecasts and reject forecasts that are inconsistent with observed events. However, it is been shown that mechanisms like tests and contracts can be manipulated by completely uninformed experts, making such screening of uninformed experts impossible. An uninformed expert would prefer to present strategic forecasts that fool a tester than to lose his reputation by refusing to be evaluated by mechanisms that informed experts would accept.

In this article we produce an original result that achieves screening of multiple potentially uninformed experts by means of pitting the different experts' predictions against each other. In the first part of the article we make a theoretical review of the problem of screening experts and we show the proof of some impossibility results in both testing settings and contract settings; our result is of the latter type.

The implications of the impossibility to screen true and false theories is very relevant to all domains were decisions are based in models. Several agencies need to anticipate future events in order to make right decisions. In economics, it is often assumed that agents know the relevant probabilities in their decision models. In reality, the determination of key relevant variables like stock prices, inflation rates, weather conditions or election outcomes are difficult to determine and the validation of experts' predictions must be carried on from a position of uncertainty about the real data generating process.

The idea that observed data can be used to validate or disprove theories is central to the scientific method. The concern that false experts can manipulate tests being uninformed can be traced back to Brier (1950), who devised a 'scoring rule' to evaluate and compare the predictability of weather forecasts from different sources. Such possibility has implications on the validity of the scientific method, where a variety of theories are presented to explain the functioning of nature and the behaviour of individuals. it is expected that the consistency of theories with observed data can reveal the true theories, however the results in this document show that in some settings this is may not be the case, and statistical tests could fail to reject false theories.

The archetypical setup can be stated as follows: Ana, a decision maker needs the probability of a sequence of binary outcomes (for example, probability of rain). Bob is a self-proclaimed expert who delivers a complete theory about such outcomes. Alice tests Bob's predictions using a test (known to Bob) based on Bob's theories and the observed outcomes.

## 1.1 Motivating Example

Bob delivers a theory $f$ about a finite sequence of binary events in $\{0,1\}^n$. This theory gives the probability of event 1 conditional on any possible finite history. For a given theory $f$, we can calculate a probability $P^f \in \Delta(\{0,1\}^n)$ (multiplying the conditional probabilities) based on which we can calculate the likelihood of possible events.

$$(0, 1, 1, 0, \dots)$$

A possible sequence of events.

Alice defines a test $T$ that assigns to each theory $f$ a rejection set $T(f) \subset \{0,1\}^n$ of observations that would be inconsistent with the theory. Since Alice wants a test that passes true theories she requires that for certain $\varepsilon > 0$ and every theory $f$

$$P^f(T(f)) < \varepsilon,$$

meaning that, if a theory is the truth, then it is rejection set has a small probability. A test with such property is said to pass the truth with probability $1 - \varepsilon$.

The requirement that tests must 'pass the truth' allows uninformed experts to pass the test. This is illustrated in the single period model, where theories are made about a single observation in $\{0,1\}$. Consider a test in this model and suppose that it 'passes the truth'. When the true distribution $P^f$ assigns equal probabilities to 0 and 1 any non-empty rejection set would have probability greater than $\frac{1}{2}$, so for $\varepsilon < \frac{1}{2}$ the rejection set $T(P^f) = \emptyset$ for that probability $P^f$. This implies that this test can be ignorantly passed by giving such uniform probability.

Now let's see how a more realistic test can be manipulated. Suppose that there are $n$ periods, and that $\frac{1}{2^n} < \varepsilon$. We define a test that, given a theory $f$, chooses the set of most unlikely observations and makes it the rejection set. Formally let $m$ be the smallest number such that $\frac{1}{2^m} < \varepsilon$. For a theory $f$, we take a set of $2^{n-m}$ histories such that the probability (under $P^f$) of this set is the lowest among all possible sets of size $2^{n-m}$, and make this the rejection set. By definition, this test passes the truth with probability $1 - \varepsilon$.[1]

Since this test has a non-empty rejection set for every theory $f$, it cannot be passed with high probability by an ignorant expert giving a single theory. Because in his worst case scenario, the rejection set of his theory has high probability. Nevertheless, the test can be ignorantly passed with high probability. Consider the set of $m-$histories, that is the set $\{0,1\}^m$. For any $m-$history $\omega$ take the theory that assigns probability 0 to all $n-$histories that coincides with $\omega$ in the first $m$ outcomes (exactly $2^{n-m}$ of them) and gives uniform probability to the other outcomes. An uninformed expert can get rejected with probability less than $\varepsilon$ by randomizing uniformly over those types of theories.

Given an observed sequence $\omega$, a theory corresponding to some $m-$history $\omega^m$ fails if the first $m$ outcomes of $\omega$ coincide with $\omega^m$, but the probability that the randomization device chooses such a theory is $\frac{1}{2^m} < \varepsilon$.

## 1.2 Literature Review

The main finding of this article is a solution to the problem of screening informed experts from uninformed ones using a contract that pits potential experts' predictions against each other. In the first section we explore results concerning a single expert setting, in here we provide the proof to the seminal result in test manipulation theory as well as other approaches to the problem involving contracts that pay for experts' information. The second section explores results concerning a multiple expert setting, where some results from single experts can be extended. Our result is presented at the end of the latter section.

A reasonable approach to evaluating a single theory is to require that it produces

[1] The probability of the less likely set of size $2^{n-m}$ is lower than the probability of such set under a uniform distribution, which would have probability $\frac{2^{n-m}}{2^n} = \frac{1}{2^m} < \varepsilon$

forecasts such that, in periods where an event was forecasted with probability close to $p$, the proportion of observed events is close to $p$. This property is known as producing calibrated forecasts. Dawid (1982) shows one of the first theoretical studies of calibration, and proves that forecasts made with the true probabilistic model are (almost surely) calibrated. However, Foster and Vohra (1998) show that an uninformed expert can produce calibrated forecasts by means of an algorithm. This result produced a rich stream of related literature and results in expert testing and test manipulation.

In the follow-up literature, several extensions of this result that go beyond calibration were produced. The main result in single expert testing being that there is no regular test (not relying on counterfactuals, and controlling for type I error) that separates informed and uninformed experts. This was proved by Sandroni (2003) in a finite expert setting and by Olszewski and Sandroni (2008) in a more general infinite horizon setting.

An alternative model to approach this problem is to screen informed and uninformed experts through contracts. Authors like Sandroni (2007) and Olszewski and Peski (2011) study contracts that would be accepted by informed experts, but rejected by uninformed experts that are averse to uncertainty. In this setting a contract is a specified money transfer occurring between the tester and the expert given his forecasts and the observed events.

Olszewski and Sandroni (2007) show that such a contract does not exist (both in a finite and infinite horizon setting) when uninformed experts are able to use mixed strategies in a zero-sum game equivalent to the contract. The uninformed expert takes advantage of the fact that informed experts always accept the contract in a way analogous to how manipulation of single expert tests is achieved. The requirement that every informed expert must accept the contract proves to give too much leeway to uninformed experts.

Varying the assumptions in the contract setting, Olszeweski and Peski (2011) try to account for the way in which the tester may benefit from knowing the expert type or the actual data generating process. They relax the requirement that good tests need to pass informed experts to focus on predictions that are useful (in a pay-off perspective) for the tester. They show the existence of a 'first-best' menu of contracts that the tester offers to the expert which ensures two things. If the expert is informed, it ensures a pay-off close to the one obtained if he had no uncertainty about the type of the expert.

If the expert is uninformed, the contract guarantees a pay-off for the expert that is only marginally lower than the one with no expert at all. This solves (in this particular sense) the problem of strategic forecasting by false experts, by means of controlling it is consequences.

Another possible solution that deals with minimizing harm of manipulation is given by Sandroni (2014). He considers a principal which has a prior distribution about the data generating process. This prior can be regarded as the odds she'd use in the absence of any expert. Sandroni shows that the principal can design a contract in which informed experts truthfully reveal their knowledge and false experts forecast the principal's prior, thus doing no harm. This solution requires only a single data point, however Sandroni points out that this contracts are unable to screen informed experts from uninformed ones, and that the no-harm result is only possible when uninformed experts do not find the prior implausible.

In the second section of this document we explore extensions of the single expert models to a multiple expert setting. The literature on multiple expert testing is diverse regarding manipulability results. The presence of multiple theories raises the possibility that, under a certain test, a theory be rejected if it is 'outperformed' by another theory. Thus, giving the tester the opportunity to at least fail some of the theories and create incentive problems for uninformed experts facing strategic uncertainty. However, the edge that the tester gets from this possibility is limited by the possibility of facing experts all of which are uninformed.

Feinberg and Stewart (2008) show that, when a true expert is present, there is a test that true experts are guaranteed to pass, and false experts will fail on all but a small (category I) set of distributions. They call this test the 'cross calibration test'. Furthermore, even when it is not known if there is true expert, a test similar to the cross calibration test cannot be manipulated over a subset of potentially true distributions [2]. Al−Najjar and Weinstein (2008) show that in a two expert setting (without loss of generality) when it is known there is a true expert, then either the two experts make very close predictions or a reputational style test will be able to pick the true expert.[3]

On the other hand Olszewski and Sandroni (2009a) give an impossibility result in

[3]They show that in an infinite horizon setting there is no way to extend the result to the case in which it is not known whether there is a true expert or not.

the multiple expert setting. Contrary to Al−Najjar and Weinstein (2008), they do not assume any prior knowledge of whether there is at least a true expert [4]. They show that multiple uninformed experts can randomize independently to produce forecasts that are likely to pass a test no matter how data unfolds, provided the test passes the truth. Independence is an extreme case that implies it holds when coordination is allowed.

To the extent of our knowledge this is the first paper to address the expert screening problem in a multiple expert setting with the use of contracts and the first result to achieve screening even when all experts are uninformed. We consider the incentives of an uninformed expert to accept a contract involving multiple experts. The uncertainty about the type of other experts allows the principal to disincentivize uninformed experts and achieve perfect screening. Moreover, we extend our results to show that there are contracts that can choose the more informed expert in a setting with partially informed experts.

# 2 Single expert setting

## 2.1 Testing under uncertainty

Testing of potential experts is an intuitive way of evaluating provided theories and a reasonable way to avoid the negative consequences of using useless predictions by uninformed experts wishing to maintain a false reputation. A tester needs to solve two problems simultaneously, the first one being that tests be passed with high probability by informed experts and the second one, that uninformed experts cannot produce strategic forecasts that can pass such tests. We say an expert is informed when his forecasts coincide with the conditional probabilities of the data generating process.

There are many tests that successfully handle the first requirement. There are some properties that correct forecasts have, based on which a tester can design tests. Correct forecasts are known to have the property that the empirical frequency of event $a$ is $p$ in periods in which $a$ was forecasted with probability $p$, also known as being calibrated (Dawid, 1982). Also, it is been shown that correct forecasts maximize the expected value of certain scoring functions like the Brier score (Brier, 1950) which can be used

[4]Their setting assumes an infinite horizon

to construct tests.

it is been mentioned that uninformed experts would be interested in passing tests using strategic forecasts to maintain a false reputation and avoid the cost of exposing their true type. However there's an important behavioural difference regarding uninformed experts that would accept being tested. Informed experts, when tested, face common risk and thus we assume they accept being tested when the probability of failing is small. Uninformed experts face uncertainty, they cannot summarise their outcomes with a probability. Throughout the literature it is assumed that uninformed experts evaluate their prospects based on the worst-case scenario, a pessimistic behavioural rule axiomatized by Gilboa and Schmeidler (1989).

In general, authors make no assumptions about the process that the outcomes follow (that it be a Markov process, or stationary process) since this is more realistic. Under this assumptions and assuming that the tester can only test forecasts based on observed data Sandroni (2003,2008) shows that uninformed experts can pass tests with high probability. Relaxing that last assumption, the tester can design elaborate tests based on unobserved data that are not manipulable, see Dekel and Feinberg (2006) and Olszewski and Sandroni (2009b) for examples of such tests and details of their construction.

## 2.2 Expert testing in finite periods

To gain some insight into how uninformed experts can pass tests in the general setting we show a proof for the finite case following Sandroni (2003). In this proof he shows how an uninformed expert can make use of the fact that tests must pass true theories to randomize strategically.

**Definition 1** (Histories)**.** A (finite) history is $h_t = (h^0, h^1, ..., h^{t-1}) \in \{h^0\} \times \{0,1\}^{t-1}$, where $h^0$ is the null history. Let $H_t = \{h^0\} \times \underset{s=1,...,t-1}{\cup} \{0,1\}^s$ be the set of all finite histories up until time $t$. Let $H_\infty = \{h^0\} \times \underset{t \geq 1}{\cup} \{0,1\}^t$ be the set of all finite histories and $\mathcal{P}(H_\infty)$ the set of all subsets of $H_\infty$. Similarly let $\mathcal{P}(H_t)$ be the power set of $H_t$.

**Definition 2** (Theory)**.** A theory is a function $f : H_t \rightarrow [0,1]$. The set of all theories is denoted by $F_t$.

A test is a function $T : F_t \to \mathcal{P}(H_t)$ that maps every theory $f$ to it is rejection set $T(f)$. That is, whenever history $h_s \in T(f)$ is observed then $f$ is rejected. Tests must have the property that if a theory is rejected at history $h_s$ then it is also rejected in all histories whose $s$ first entries coincide with $h_s$. Such histories are known as the extensions of $h_s$. A theory $f$ returns the probability of a binary event at each period $t$ taking as input a finite history $h_t$ . Given a theory $f$ we can define a corresponding probability $P^f$ on $\mathcal{P}(H_t)$ in a natural way using Bayes' rule. In order for the test to control for type I errors we require that the probability of the rejection set of any $f$ (under $P^F$) be less than $\varepsilon$. This is known as "passing the truth" with probability $1 - \varepsilon$ and it's summarized in the following condition:

$$P^f\left(T(f)\right) < \varepsilon.$$

Uninformed experts may make use of randomization over the space of possible theories to pass a test with high probability. This results in the following definition of manipulation:

**Definition 3** (Manipulability)**.** Let $F$ be a space of possible theories. A test $T$ can be ignorantly passed with probability $1 - \varepsilon$, or it is $1 - \varepsilon$ manipulable if there exists a random generator of theories $\xi \in \Delta(F)$ such that for every finite history $h_t \in H_t$:

$$\xi(\{f \in F : h_t \in T(f)\}) \leq \varepsilon.$$ [5]

The random generator $\xi$ may depend on the test $T$, but not on any knowledge of the data-generating process. If a test can be ignorantly passed, the expert can randomly select theories that, with probability $1 - \varepsilon$ (according to his randomization), will not be rejected, no matter what data are observed in the future.

Sandroni (2003) shows how uninformed experts can manipulate tests by means of such random generators of theories. In the proof, he reduces the problem to a zero-sum game between the expert and nature (the test is given). Where the expert's strategies are a finite subset of the theories $h_t$ and nature's are the observed outcomes.

**Proposition 1** (Sandroni (2003))**.** Given $t \in \mathbb{N}$ and $\varepsilon > 0$ , if a test $T : \Delta(H_t) \to \mathcal{P}(H_t)$ passes the truth with probability $1 - \varepsilon$, then it can be $1 - \varepsilon$ manipulable.

[5] In the infinite horizon setting this will require that the set $\{f \in F : h_t \in T(f)\}$ be measurable.

*Proof.* Since $H_t$ is finite then so is $\mathcal{P}(H_t)$ and we can find a find a finite $Y \subset \Delta(H_t)$ such that $T(Y) = T(\Delta(H_t))$. Now consider the following zero sum game between the expert and nature: The expert's strategy space is $Y$ and nature's strategy space is $H_t$. The expert's payoff is $\Pi : Y \times H_t \to \{0, 1\}$ where $\Pi(f, h_t) = 1$ if and only if $h_t \notin T(f)$ and zero otherwise. By Von-Neumann's minimax theorem:

$$\max_{\xi \in \Delta(Y)} \min_{\sigma \in \Delta(H_t)} \Pi(\xi, \sigma) = \min_{\sigma \in \Delta(H_t)} \max_{\xi \in \Delta(Y)} \Pi(\xi, \sigma).$$

1. We exploit the right hand side of the inequality: Given $\sigma \in \Delta(H_t)$, let $p^* \in Y$ be such that $T(p^*) = T(\sigma)$, then:

$$\begin{aligned}\Pi(f^*, \sigma) = \int_{H_t} \Pi(f^*, h_t) d\sigma(h_t) = \int_{\{h_t \in H_t : h_t \notin T(f^*)\}} 1 d(\sigma(w)) \\ = \sigma(\{h_t \in H_t : h_t \notin T(f^*)\}) \geq 1 - \varepsilon,\end{aligned}$$

   hence:

$$\max_{f \in Y} \Pi(f^*, \sigma) \geq 1 - \varepsilon.$$

2. Since the above is true for all $\sigma$ then exploiting the left hand side we have:

$$\min_{\sigma \in \Delta(H_t)} \max_{\xi \in \Delta(Y)} \Pi(\xi, \sigma) \geq 1 - \varepsilon$$

$$\max_{\xi \in \Delta(Y)} \min_{\sigma \in \Delta(H_t)} \Pi(\xi, \sigma) \geq 1 - \varepsilon.$$

   Therefore, for some $\xi_* \in \Delta(Y)$

$$\min_{\sigma \in \Delta(H_t)} \Pi(\xi^*, \sigma) \geq 1 - \varepsilon$$

   In particular, for all $h_t$

$$\Pi(\xi^*, h_t) \geq 1 - \varepsilon$$

□

We notice that in the proof of Proposition 1, nature gains no advantage of randomizing over its strategy space $H_t$, since this is offset by the fact that the test passes the truth. This gives the expert a pure strategy that is best-response for every mixed strategy nature has.

## 2.3 Manipulability of infinite horizon tests

In the infinite horizon the definitions of finite history and theory extend naturally. The difference is that now theories can produce forecasts after finite histories of arbitrary length. This means we extend the domain of theories to $H_\infty$. The set of all theories is denoted by $F$.

Given a theory $f : H_\infty \to [0,1]$ we wish to construct an equivalent probability $P^f$ for every element in $H_\infty$. Like in the finite case we could compute the probability of one finite history $h_m = (h^1, h^2, \ldots, h^{m-1})$ by means of Bayes' rule as:

$$\prod_{t=1}^{m} f(h_t)^{h^t}(1 - f(h_t))^{1-h^t},$$

where $h_t = h_m|t$ is the restriction of $h_m$ to its first $t$ coordinates, for $t \leq m$.

However, unlike the finite case, such function $P^f$ is not a probability distribution over $H_\infty$. Nonetheless, such function is a probability when we consider $H_{infty}$ as embedded in the set of infinite histories $\{0,1\}^\infty$. In such embedding we associate the finite history $h_m$ with the set of all infinite histories that begin with $h_m$. This set is denoted $K(h_m)$ and it's called the cylinder with base $h_m$. So when we refer to the probability $P(h_m)$ we will be referring to $P\left(K(h_m)\right)$. [6]

This embedding also allows for a conversing correspondence. Namely for every probability distribution $P$ over $\{0,1\}^\infty$ we can define a corresponding theory using Bayes' rule:

$$f_P(h_t) = \frac{P(C(h_t,1))}{P(C(h_t))} \text{ if } P(C(h_t)) > 0.$$

[6] The set $\{0,1\}^\infty$ is endowed with the product topology, in which its basic sets are precisely the cylinders $K(h_t)$. This can be made a measure space by endowing it with the $\sigma-$algebra generated by all finite unions of cylinders. The probability of a cylinder is naturally associated with the probability of it is finite history, defining a bijection between $\Delta(H_\infty)$ and $\Delta(\{0,1\}^\infty)$

Once we've carefully defined $P^f$ the definitions of tests and of "passing the truth" are the natural extensions of the definitions for the finite case.

Two theories $f$ and $g$ are said to be equivalent up to period $m$ if for every finite history $h_t$ such that $t \leq m$ we have $f(h_t) = g(h_t)$.[7] Olszewski and Sandroni (2009b) and Dekel and Feinberg (2006) show tests that cannot be manipulable, but they make use of forecasts of unobserved events. However, we are interested in the scenario where a tester does not have such counterfactual scenarios at his disposal. For example when experts only give forecasts of the next period. If such a test were to reject a certain theory at period $m$, it would have to reject any other theory equivalent up to period $m$. We call these tests 'future independent'.

**Definition 4** (Future independent tests)**.** A test $T$ is future independent if given two theories $f$ and $g$ equivalent until period $m$, $s_t \in T(f), t < m$, implies $s_t \in T(g)$.

**Definition 5** (Regular Test)**.** A test $T$ is a regular $\varepsilon$ test if it is future independent and passes the truth with probability $1 - \varepsilon$.

As before, the expert is allowed to randomize by choosing a probability function $\xi \in \Delta(F)$ called a random generator of theories. Note that we need to equip $F$ with a $\sigma$-algebra and we will need the sets $\{f \in F : s_t \in T(f)\}$ to be measurable. The definition of manipulation of tests is the same as in the finite case.

**Proposition 2** (Sandroni, 2008)**.** For every $\varepsilon > 0$ and $\delta \in (0, 1 - \varepsilon]$. Any regular test (future independent that does not reject the data generating process with probability $1 - \varepsilon$) can be ignorantly passed with probability $1 - \varepsilon - \delta$.

Sandroni proves proposition 2 following the same intuition as in the finite case, studying a zero-sum game between the expert and nature. The body of the proof is mostly reducing the strategy space of Bob so that Fan's minimax theorem, a generalization of Von Neumann's minimax theorem, can be applied. Sandroni begins by stating two lemmas which allow for such reduction into compact and Hausdorff space, and allow the payoff function to be lower semi continuous.

*Proof.*

[7] This concept is very important when considering the weak-* topology over $F$. The set of all theories equivalent up to period $m$ to a given theory $f$ is an open set in this topology.

**Lemma 1.** Given an open subset $U \subset X$ of a compact metric space $X$.The function $H : \Delta(X) \to [0, 1]$ dened by

$$H(P) = P(U), \tag{1}$$

is lower semi continuous

*Proof.* The proof is pretty straightforward from the definition of lower semi continuity, for details see Dudley (1989), proposition 11.1.1 (b) □

**Definition 6.** Let $\delta > 0$. A set $F' \subset F$ is called $\delta-$subdense in $F$ if for every theory $g \in F$ there exists $f \in F'$ such that

$$\sup_{A \in \Im} |P^f(A) - P^g(A)| < \delta. \tag{2}$$

The idea is to construct a compact set which is $\delta-$subdense in $F$. For the sequence $\gamma$ of positive numbers $\{\gamma_t\}_{t=0}^{\infty}$ Sandroni constructs the sequence $R$ of finite sets $\{R_t\}_{t=0}^{\infty}$ such that $R_t \subset (0, 1)$, $t \in \mathbb{N}$ and

$$\forall_{x \in [0,1]} \exists_{r \in R_t} : \ |x - r| < \gamma_t \tag{3}$$

For such $R$ he constructs the set $\bar{F} = \{f \in F : \forall_{t \in \mathbb{N}} \forall_{s_t \in \{0,1\}^t} f(s_t) \in R_t\}$.

**Lemma 2.** For every $\delta > 0$ there exists a sequence of positive numbers $\gamma$ such that $\bar{F}$ is $\delta-$subdense in F .

*Proof.* This result is constructive, for the technical details see Sandroni and Olszewzki (2008), lemma 2. □

The need for lemma 1 and lemma 2 is that Fan's theorem requires the space of strategies of the expert to be compact and Nature's payoff function to be lower semi continuous. That we are considering only future independent tests guarantees that the second condition is met. Lemma 2 is used to create a pure strategy space for the expert that is compact when endowed with the product of discrete topologies, so that it is mixed strategy space is compact when using the weak$-*$ topology. To complete the proof, Fan's theorem is used:

**Theorem 1** (Fan (1953))**.** Let $X$ be a compact Hausdorff space, which is a convex subset of a linear space; and let $Y$ be a convex subset of linear space (not necessarily topologized). Let $G$ be a real-valued function on $X \times Y$ such that for every $y \in Y$, $G(x, y)$ is lower semi continuous with respect to $x$. If $G$ is also convex in respect to $x$ and concave in respect to $y$ then:

$$\min_{x \in X} \sup_{y \in Y} G(x, y) = \sup_{y \in Y} \min_{x \in X} G(x, y).$$

Make $X = \Delta(\bar{F})$. And $Y \subset \Delta(\Omega)$ the subset of theories with finite support, so that an element $P \in Y$ is described as a finite set of paths $\{\omega^1, \ldots, \omega^n\}$ with respective weights $\{\pi_1, \ldots, \pi_n\}$. We define $G : X \times Y \to \mathbb{R}$ as

$$G(\zeta, P) := \sum_{i=1}^{n} \pi_i \zeta \left( \{ f \in \bar{F} : \exists_{t \in N} \omega_t^i \in T(f) \} \right).$$

That is, the probability that the expert's theory would be rejected if he chooses $\zeta$ from set $X$ and nature chooses $P$ from set $Y$.

**Lemma 3.** *$G$ is a lower semicontinous function of $\zeta$.*

*Proof.* Note that the set $\{f \in \bar{F} : \omega_t^i \in T(f)\}$ is open. This is because the test $T$ is future independent, and therefore if a theory $f$ is in this set, then all theories equivalent up to period $t$ (an open set) are also in the set. This implies that the set is open in $\bar{F}$ with subspace topology (that is the product of discrete topologies). Therefore the set

$$\{f \in \bar{F} : \exists_{t \in N} \omega_t^i \in T(f)\} = \cup_{t \in \mathbb{N}} \{f \in \bar{F} : \omega_t^i \in T(f)\}$$

is also open, and by lemma 1 we get the desired result. The function $G$ is lower semi continuous as it is the weighted average of lower semi continuous functions of the form:

$$H(\zeta, \omega) = \zeta \left( \{ f \in \bar{F} : \exists_{t \in N} \omega_t^i \in T(f) \} \right).$$

□

Now it can checked that the conditions of Fan's theorem hold. it is already been proven that $G$ is lower semi continuous with respect to $\zeta$. Note that $G$ is concave on $Y$ and convex on $X$ since it is linear in both variables. By Banach-Alaoglu's theorem it is obtained that $X$ is a compact space in weak$-*$ topology; it is a metric space and therefore it is Hausdorff. Thus, by Fan's theorem,

$$\min_{\zeta \in X} \sup_{P \in Y} G(\zeta, P) = \sup_{P \in Y} \min_{\zeta \in X} G(\zeta, P).$$

We now note that the right hand of the equation falls below $\varepsilon + \delta$, because the test $T$ does not reject the truth with probability $1 - \varepsilon$. So for a given $P \in Y$ we have a theory $g \in \bar{F}$ that generates a measure that is very close. Namely, that for every set $A \in \Im$ $|P(A) - P^g(A)| < \delta$ (lemma 2). So in the case were $\xi$ is a random generator of theories such that $\xi(\{g\}) = 1$ we get:

$$G(\xi, P) = P(T(g)) \leq P^g(T(g)) + |P(T(g)) - P^g(T(g))| < \varepsilon + \delta$$

Therefore there must exist a random generator of theories such that:

$$G(\zeta, P) \leq \varepsilon + \delta,$$

for every $P \in Y$, since this includes the elements of $Y$ that concentrate mass on a single path, this concludes the proof.

□

We now move on to explore results on models with a different approach to the problem of screening informed experts from uninformed ones. Rather than constructing tests that informed experts pass and uninformed experts fail, we study contracts that pay for an expert's forecasts depending on the observed outcome. The following results inquire about contracts that only an informed expert would accept.

## 2.4 Manipulation of contracts

From a decision theory standpoint, information can be seen as just a tool for making better decisions. So we can model the interaction between a decision maker and a potential expert focusing on their payoffs and not explicitly in the screening problem. We assume the decision maker is a principal willing to pay large amounts of money for relevant probabilities about an outcome. She can offer a potential expert a contract at period 0. If he accepts the contract he must announce a probability distribution before any data is observed, and, when the data unfolds, transfers occur between the principal and the expert.

Contracts can be designed so that their transfers depend on the predictive performance of the theories presented by experts, so that an informed expert can receive a positive payoff but an uninformed expert (if accepting the contract) would receive negative payoffs. A contract like that would perfectly screen informed and uninformed experts. it is assumed that uninformed experts are uncertainty-averse and they only accept contracts when they can profit in the worst case scenario. We will follow the work by Olszewsky and Sandroni (2008) and show the non-existence of a contract that screens between informed and uninformed experts for the single period case and its extension to an infinite period model.

**Definition 7.** A contract is a function $C : \{0,1\} \times [0,1] \to \mathbb{R}$ specifying the transfer made to the expert when he announces probability $f$ and the outcome is $i$.

A contract specifies wealth transfers at period 0 depending on the outcome of nature and the predicted probabilities given by the expert. As in the previous model, the expert might be informed or uninformed. If informed, he knows the true probability $f$ of outcome 1 and faces risk in his payoff. If uninformed, he knows nothing about the probability of 1 and under uncertainty we assume that his behaviour is guided by the payoff in his worst case scenario. We assume that the expert's wealth at period 0 is $w > 0$ and if he accepts the contract, he's wealth at period 1 is $w + c(i, f)$.[8]

The following assumptions are made for the sake of simplicity:

1. An informed expert is assumed to always reveal his knowledge truthfully. This simplifies the problem, as it removes the possibility that he misrepresents the truth to get a higher payoff.
2. There exists $\eta > 0$ such that for every $f \in [0,1]$ and $i \in \{0,1\}$, if $c(i,f) < 0$ then $c(i,f) < \eta$. That is, there is a minimum cost (maximum negative transfer) to Bob if he must pay Alice.
3. For every $f \in [0,1]$ there is an outcome $i \in \{0,1\}$ such that $c(i,f) < 0$ and $c(j,f) \geq 0$ for $j \neq i$. Meaning that for every theory there is a favourable outcome and an unfavourable outcome.

[8]We could dispense of this assumption, but it is convenient for the expert to be solvent in case he needs to transfer money to the expert.

**Definition 8.** The contract $C$ is accepted by an informed agent when

$$\forall f \in [0,1] \quad fu(w + c(1,f)) + (1-f)u(w + c(0,f)) \geq u(w)$$

we say he strongly accepts the contract if there exists $\delta > 0$ such that:

$$\forall f \in [0,1] \quad fu(w + c(1,f)) + (1-f)u(w + c(0,f)) \geq u(w) + \delta$$

Where $u$ is a strictly increasing utility function.

Under assumptions 2 and 3 strongly accepting a contract is equivalent to accepting a contract. This follows from the continuity of the payoff function, for details see lemma 1 of Olszewski and Sandroni (2007).

The uninformed expert knows nothing about the true probability, but he's allowed to randomize as in the previous model. In this case, the uninformed expert will accept the contract if he has a randomization that will give him a non negative pay-off in his worst case scenario. We will continue to use Gilboa and Schmeidler's (1989) model for decision under uncertainty.

**Definition 9.** Let $\xi \in \Lambda[0,1]$ be a probability measure over $\Delta(\{0,1\}) = [0,1]$ with finite support; where $\xi(f_i) = \gamma_i$ for $f_1, \dots, f_n, \gamma_1, \dots, \gamma_n \in [0,1]$. Then in realization $i$ the expert gets expected utility:

$$\xi(i) \equiv \sum_{j=1}^{n} \gamma_j u\left(w + c(i, f_j)\right)$$

And he's said to accept the contract $c$ if there exists $\bar{\xi} \in \Lambda[0,1]$ such that

$$\min_{f' \in [0,1]} f'\bar{\xi}(1) + (1-f')\bar{\xi}(0) \geq u(w). \tag{4}$$

**Proposition 3** (Olszewsky and Sandroni (2008))**.** If an informed expert strongly accepts a contract $c$ then an uninformed expert also accepts contract $c$.

*Proof.* The proof is very similar to the result in Proposition 1, since the contract specified can be modelled as a zero-sum game between the expert and nature. When nature chooses a mixed strategy $f' \in \Delta(\{0,1\})$ and the expert chooses $\xi \in \Lambda[0,1]$ then the expert's payoff is:

$$f'\xi(1) + (1-f')\xi(0).$$

For every mixed strategy $f'$ of Nature, there's a strategy for the expert (selecting $f'$ with full mass) that yields a payoff of:

$$f'u(w + c(1, f')) + (1 - f')u(w + c(0, f')) \geq u(w) + \delta.$$

The inequality comes from the fact that an informed expert would strongly accept $c$. Therefore, if the conditions of Fan's minimax theorem are met, there's a strategy $\bar{\xi}$ that ensures the expert a payoff greater than $u(w) + \delta$ no matter which strategy nature chooses.

The payoff function is linear in both $f'$ and $\xi$, and also it is continuous on $f'$ which belongs to the compact Hausdorff space $[0, 1]$ so the conditions of Fan's theorem are met and the result follows. □

The results in this single period setting can be generalized to partially informed experts. That is, experts that know the true theory is in a set $B_\varepsilon(f^*) := \{f \in [0,1] : |f - f^*| < \varepsilon\}$ for $\varepsilon > 0$. Since the payoff function is continuous, the definition of properly accepting a contract extends nicely to this case and the proof is nearly identical.

This model can also be generalized to infinite period models when the utility is discounted according to a factor $\beta < 1$. For this case, the expert gives probabilities on $\Omega = \{0,1\}^\infty$ and uninformed experts may randomize over $\Lambda(\Delta(\Omega))$ (randomizations with finite support).

The definition of a contract changes slightly to a function $C : H_\infty \times F \to \mathbb{R}$ that specifies transfers $C(h_t, f)$ at every period $t$. The extension of Proposition 3 to the infinite horizon will make use of Fan's (1953) theorem, but to fulfil its conditions we make sure that the payoff function is continuous. This we achieve by requiring that for every $f \in F$ the transfer $C(\cdot, f)$ is bounded and that the expert has an endogenous endowment $w_t$ at each period that is uniformly bounded.

In this scenario the criterion for accepting or rejecting a contract will depend on the present value (using the discount value $\beta$) of the utility relative to not taking the contract. The utility difference with or without contract is:

$$U(h,f) = \sum_{t=0,\ldots,\infty} \beta^t (u(w(h|t) + C(h|t,f)) - u(w(h|t))$$

**Definition 10.** The contract is $C$ strongly accepted by the expert if when informed, for some $\delta > 0$

$$\forall f \in F, E_f[U(\cdot, f)] \geq \delta$$

and when uninformed, for some $\delta > 0$ there is $\bar{\xi} \in \Lambda(\Delta(\Omega))$ such that:

$$\forall h \in H, E_{\xi^{C_\delta}}[U(h,\cdot)] \geq \delta$$

Olszewski and Sandroni prove the following extension of their result:

**Proposition 4** (Olszewski and Sandroni (2007))**.** If $C$ is strongly accepted by an informed expert then it is strongly accepted by an uninformed expert.

*Proof.* This proof is almost exactly the same as in the single period model. The difference is that they need to prove that $U(h,f)$ is continous on $h$ in order to use Fan's (1953) theorem. See proposition 2 in Olszewski and Sandroni (2007) for details on this proof. □

These results have a key resemblance with Proposition 1. When reduced to a zero-sum game the expert has a strong edge over nature; For every mixed strategy of nature the expert has a best response in the set of pure strategies. This results in nature not gaining any advantage from randomizing. Such edge comes from the condition that tests "pass the truth" and equivalently that a contract always be accepted by informed experts. Both condition prove to be too restrictive and make it impossible to screen uninformed experts. However, with contracts it is possible to give incentives to potential experts to obtain a second-best result; a contract in which uninformed experts do no harm and informed experts honestly reveal their knowledge.

## 2.5 A simple 'No Harm' contract model

Although the previous results show that an uninformed decision maker cannot know the type of an expert by means of a contract, the tester can minimize the harm that could come from made up theories from uninformed experts. Let's assume that the decision maker has a prior $f^*$. This prior can be thought of as the odds on which shed would base its decisions in the absence of an expert's forecast. If we assume that it is very costly for the decision maker to use false odds different to her prior, then she can achieve a second-best result; a contract in which informed experts honestly reveal their knowledge and uninformed experts forecast the prior $f^*$.

This positive result allows the decision maker to benefit from the knowledge of informed experts without the dangers derived from harmful strategic odds provided by uninformed experts. However, this result requires a tester that is willing to pay amounts of money to any false expert that takes the contract and that the tester's prior be in the set of 'conceivable' theories of the uninformed expert. Such set will be denoted $\Theta_B$ and we will suppose it is convex and closed.[9]

**Proposition 5** (Sandroni(2014))**.** Let $f^*$ be a prior by the tester about the data generating process. If $f^* \in \Theta_B$ then there exists a **contract** $\widehat{C}$ such that true experts choose to reveal the truth and false experts choose to forecast $f^*$

*Proof.* Let $\bar{C} : \Delta(S) \times S \longrightarrow \mathbb{R}$ be a contract based on a proper scoring rule (E.g. Brier(1954)). thus:

$$\mathbb{E}^f(\bar{C}(f, \cdot)) > \mathbb{E}^f(\bar{C}(f', \cdot) \text{ For } f \neq f' \tag{5}$$

Let $\varepsilon > 0$, we construct the following contract: $\hat{C}(f, s) = \bar{C}(f, s) - \bar{C}(f^*, s) + \varepsilon$

Note that this new score is still a proper scoring rule, meaning that

$$\mathbb{E}^f(\hat{C}(f, \cdot)) > \mathbb{E}^f(\hat{C}(f', \cdot))..$$

[9]We cannot relax this assumption, or else the uninformed expert might accept the contract and randomize over harmful odds. If $\Theta_B$ is not a closed set, then we cannot guarantee that the minimum of a payoff function is reached inside the set.

This follows from the fact that $\bar{C}$ is a proper scoring rule. So the informed expert has incentives to reveal the truth.

The uninformed expert, according to the MinMax criterion, must maximize:

$$\max_{\zeta \in \Delta(\Delta(\Omega))} \min_{f \in \Theta_B} \int_{\Delta(\Omega)} \mathbb{E}^f \hat{C}(f', \cdot) d\zeta(f') \tag{6}$$

We now show that $\zeta(f^*) = 1$. Note that $\hat{C}(f^*, s) = \varepsilon \quad \forall s$. So if full mass is given to $f^*$ we have

$$\min_{f \in \Theta_B} \int_{\Delta(\Omega)} \mathbb{E}^f \hat{C}(f', \cdot) d\zeta(f') = \min_{f \in \Theta_B} \mathbb{E}^f \hat{C}(f^*, \cdot) = \varepsilon \tag{7}$$

Now it remains to prove that such choosing of $\zeta$ maximizes the uninformed expert's pay-off. For this note that, since $f^* \in \Theta_B$ then:

$$\min_{f \in \Theta_B} \int_{\Delta(\Omega)} \mathbb{E}^f \hat{C}(f', \cdot) d\zeta(f') < \int_{\Delta(\Omega)} \mathbb{E}^{f^*} \hat{C}(f', \cdot) d\zeta(f'), \tag{8}$$

but since $\hat{C}$ is a proper scoring rule,

$$\varepsilon = \mathbb{E}^{f^*} \hat{C}(f^*, \cdot) > \mathbb{E}^{f^*} \hat{C}(f', \cdot) \qquad \text{for } f' \neq f^*,$$

so assigning positive mass to any point other than $f^*$ is suboptimal. □

This contract gives a result that contrasts with Propositions 2 and 3 in the sense that it is a positive result and (partial) screening is achieved. However, there are some caveats. First, it depends on which odds are considered non-harmful by the principal and it requires that $f^*$ is conceivable for uninformed experts, since the principal can't be sure about this her worst case scenario is inconvenient. Second, since it does not penalize uninformed experts there is no way to disincentivize them from taking the informed expert's job. Clearly a contract that ensures a payment to anyone is inconvenient.

We propose a slight variation of this result in order to attend the second concern. We relax the requirement that the contract needs to be accepted by informed experts and assume that the decision maker is only willing to pay for predictions that are different enough from her prior $f^*$. This is a more realistic assumption and achieves the same result, but without giving incentives to uninformed experts to take the job.

**Proposition 6.** Let $\delta > 0$ and let $f^*$ be a prior by the tester about the data generating process. Let $\Theta_B$ be the set of possible probability distributions of the sequence of events that Bob finds plausible. If $f^* \in \Theta_B$ then there exists a **contract** $\widehat{C}$ such that true experts whose prediction is at a distance greater than $\delta$ from the prior choose to reveal the truth and false experts reject the contract.

*Proof.* Take a continuous strictly proper scoring rule $B$ (e.g. Brier (1950)) then we know that there's an $\eta > 0$ such that if $f \in \Delta(S)$ is such that $||f - f^*|| < \delta$ then

$$E^f(B(\mathbf{f}, s)) - E^f(B(\mathbf{f}^*, s)) \geq \eta,$$

this follows from:

1. $E^f(B(\mathbf{f}, s)) - E^f(B(\mathbf{f}^*, s))$ is a continuous function of $f$, since it is the weighted average of continuous functions.
2. The set $\{f \in \Delta(S) : ||f - f^*|| \geq \delta\}$ is a closed subset of the compact set $\Delta(S)$ (since $S$ is finite), and therefore it attains a minimum $\eta$.
3. $\eta > 0$ since $B$ is a strictly proper scoring rule,

this motivates the contract $\widehat{C}(f, s) = B(\mathbf{f}, s)) - B(\mathbf{f}^*, s) - \eta$.

It immediately follows that the contract would be accepted by true experts when the true distribution is bounded away from $f^*$. To see that uninformed experts wouldn't accept this contract is straightforward, because their worst case scenario payoff is bounded by

$$E^{f^*}(B(\mathbf{f}, s)) - E^{f^*}(B(\mathbf{f}^*, s)) - \eta < -\eta$$

the inequality coming from the fact that $B$ is a proper scoring rule. □

From the principal's perspective this modified contract achieves the same as the contract in Proposition 5. Namely, when the true odds are different from the prior $f^*$ the principal obtains them (because true experts have incentives to reveal them), and when the expert is uninformed or the true odds are $f^*$ then she uses $f^*$ (because uninformed experts reject the contract). Nonetheless this contract is better for the principal since now she doesn't have to pay uninformed experts for useless predictions and she desincentivizes them from taking the true expert's job. This doesn't screen uninformed experts from informed experts completely, but it screens valuable experts from the rest.

# 3 Testing Multiple Experts

## 3.1 Comparing multiple theories

Some researchers like Al-Najjar and Weinstein (2008), Feinberg and Stewart (2008) or Olszewski and Sandroni (2008) show that the screening of potentially uninformed experts can be highly influenced by the fact that a single expert is tested in isolation, or multiple experts are tested at the same time. However, they point out that under certain assumptions the impossibility result of Proposition 2 extends to the multiple expert setting.

The main advantage for a tester in this multiple expert setting is that different uninformed experts who act independently may produce different theories. In the face of differing theories, the tester can observe that one outperforms the other (e.g. is more likely than the other). This would lead to the rejection of at least one theory.

This hurdle is difficult to overcome and, for example, makes manipulation very difficult when one of the experts is informed.[10] Al-Najjar and Weinstein (2008) provide a reputation test in a finite period scenario that chooses the right expert (it can be inconclusive) and has the following property:

**Proposition 7** (Al-Najjar and Weinstein (2008))**.** If expert $i$ is informed and truthful, then for every $\varepsilon > 0$, there is an integer $k$ such that for all integers $n$, true data-generating processes $f$, and random generators of theories $\xi^j$ of expert $j \neq i$, the probability of the event that:

- $T$ selects expert $i$.
- The forecasts of outcome 1 made by the two forecasters differ by at most $\varepsilon$ in all but $k$ periods,

is no lower than $1 - \varepsilon$.

This result is powerful, taking into consideration that $k$ does not depend on $n$. The only room for manipulation would be to satisfy the second condition, but this is highly

[10] In general, it is not realistic to expect that the tester knows this kind of information.

unlikely for high values of $n$ compared to $k$.[11]

## 3.2 Manipulability in a multiple experts model

Olszewski and Sandroni (2009a) provide an extension of proposition 2 to the multiple expert setting. Now tests are allowed to reject experts based on both the observed outcome and the predictions of other experts. Olszewski and Sandroni (2009a) do not allow the experts to collude. If they did then they could produce identical forecasts and behave like a single expert, and by Proposition 2 we know that single experts can manipulate regular tests. Therefore, if the uninformed experts are going to use mixed strategies, it is required that these be independent random variables. We extend the definition of test to multiple experts in a natural way, by extending its domain to pairs of theories (corresponding to each of the experts) and returning a pair of rejection sets. Since the rejection sets of both experts depend on both theories, it is said that they 'compare' the theories and in the literature they are often referred to as 'comparative tests'. Formally, test will now extend to be functions of the form

$$T : F \times F \to \mathcal{P}(H_\infty) \times \mathcal{P}(H_\infty)$$

Where $F$ is the set of all possible theories. For simplicity we will use $T_i(f_1, f_2)$ to refer to the $i$ coordinate of $T(f_1, f_2)$, and we shall call this the rejection set of theory $f_i$ conditional on $f_{-i}$. We require that if $h_m \in T_i(f_1, f_2)$ then all of its extensions are also in $T_i(f_1, f_2)$.

We extend the definition of "passing the true data generating process" or "passing the truth" with the caveat that now the other expert's theory must be taken into account. Formally, we say that a test $T$ "passes the truth" with probability $\varepsilon$ when for every $f_1$ and $f_2 \in F$ we have

$$P^{f_1}(T_1(f_1, f_2)) \leq \varepsilon$$

$$P^{f_2}(T_2(f_1, f_2)) \leq \varepsilon,$$

[11]This test can't detect false experts when both experts are uninformed.

so passing the true theory does not depend on the other expert's announced theory.

The concept of being future independent also extends naturally from the single expert model. In the sense that a test $T$ is future independent if when theories $f_1$ and $f_2$ are equivalent up to period $m$ to theories $g_1$ and $g_2$ respectively, then $h_t \in T_i(f_1, f_2)$ implies $h_t \in T_i(g_1, g_2)$ for every $h_t \in H_m$.

The definition of manipulability is similar but has some important differences. It requires that, for every finite history $h_t$ the joint probability of rejection for a pair of random generators of theories be small. We note however that it does not require that the probability of rejection under one random generator of theories conditional on any strategy by the other expert be small, this means that a manipulable test could be failed by a false expert if the other expert does not use the right strategy (particularly if he says the truth). The definition is as follows:

**Definition 11** (Manipulability)**.** A test $T$ can be ignorantly passed, or it is manipulable with probability $1-\epsilon$, if there exists a pair of independent random generators of theories $\xi_1, \xi_2$ such that for both $i = 1$ and $i = 2$ and for every finite history $h_t$:

$$\xi_1 \times \xi_2(f_1, f_2 \in F \times F : h_t \in T_i(f_1, f_2)) \leq \epsilon \tag{9}$$

**Proposition 8** (Olszewski and Sandroni (2009a))**.** For every $\epsilon > 0$ and $\delta \in (0, 1-\epsilon]$. Any regular test (future independent test that does not reject the data generating process with probability $1-\epsilon$) can be ignorantly passed with probability $1-\epsilon-\delta$.

*Proof.* The intuition of the proof is similar to that of Sandroni (2008). Given a randomizing $\xi_1$ strategy by expert 1, this defines a single expert test for expert 2 which is also regular. Therefore by proposition 1 in Sandroni (2008), proved in this work, there exists a randomizing strategy for expert 2, namely $\xi_2$, which ensures him an expected pay-off of at least $1-\epsilon-\delta$.

This implies that for every randomizing strategy $\xi_1$ of expert 1, there exists a randomizing strategy $\xi_2$ of expert 2 that ensures him an expected pay-off arbitrarily close to $1-\varepsilon$ regardless of how nature unfolds. Conversely for every randomizing strategy $\xi_2$ of expert 2, there's a randomizing strategy $\xi_1$ for expert 1 that ensures him a pay-off arbitrarily close to $1-\varepsilon$ regardless of how nature unfolds. Applying

Glicksberg−Kakutani's fixed point theorem, one can show the existence of a pair of independent random generators of theories $(\xi_1, \xi_2)$ that ensure each expert an expected pay-off arbitrarily close to $1 - \varepsilon$ no matter how nature unfolds. □

This result shows that every test can be simultaneously manipulable by false experts. In the appendix we present an example (due to Olszewski and Sandroni (2009a)) showing how difficult this might be for a false expert. Randomizing independently raises the possibility of theories being very different and thus one being more likely than the other, and therefore outperforming it under certain tests. However, thanks to Proposition 8 we know that such paradoxical pair of randomization devices that are both independent and produce close results is possible.

Something to remark about Proposition 8 is that the definition of manipulation used in its model does not imply that an uninformed expert can pass a test with high probability. A test is manipulable if two uninformed experts jointly use a specific pair of mixed strategies. But it's possible for an uninformed expert to have a high probability of failing the test if the other expert doesn't use the specific randomization device that is required to manipulate it, for example, if he is an informed expert who reveals the truth.

This introduces strategic uncertainty, another source of uncertainty coming from ignoring the other expert's type. An uncertainty averse expert would evaluate his prospects using a worst case scenario analysis and thus could discard this manipulation strategy if his worst case scenario is when the other expert is informed. However, this is not enough for false experts to reject the test [12]. Rejecting the test brings about the same reputational cost than failing the test. Even if when facing an informed expert there's no strategy for uninformed experts that passes the test with high probability, the slightest chance of facing an uninformed expert and manipulating the test is enough incentive to use the randomization device that Proposition 8 proposes. So the tester could still see a scenario where rational uninformed experts pass her test.

Nonetheless, strategic uncertainty is a very influencing factor for uncertainty averse experts. We will move on from the testing approach to screening experts and present an original result in a contract setting. We capitalize on this strategic uncertainty and show that the slightest chance of facing an informed expert is enough to disincentivize

[12] Thanks to lvaro Sandroni for pointing this out.

an uninformed expert from taking some contracts that informed experts would take.

## 3.3 Screening multiple experts with contracts

Using simple modifications of the Brier score (1950) we can devise a contract that gives payoffs proportional to the performance of an expert's theory relative to that of a rival expert. It is clear that, in this setting, the worst case scenario for a false expert would be that the rival expert's theory performs very well, for instance, if it is the true theory. In such worst case scenario, the only way to profit from this contract is to forecast theories close to the truth, and we will show this is impossible if the expert is completely uninformed.

Let $S$ be a finite set of states (e.g. a set of possible finite histories). Let $\Delta(S)$ be the set of probability distributions over $S$. Two experts, referred to as expert 1 and expert 2, deliver probabilistic forecasts $f_1$ and $f_2 \in \Delta(S)$ to a principal.

The principal designs a contract that specifies money transfers between her and each expert to elicit information. A contract is a payoff function $C : \Delta(S) \times \Delta(S) \times S \to \mathbb{R}$ whose value depends on the announced odds and the observed state. If any expert rejects the contract his payoff is 0. [13] For simplicity I assume that the contracts given to each expert are symmetrical, and I'll focus on the behaviour of expert 1, who is offered a contract $C_1$, the behaviour of expert 2 will therefore be symmetrical as well. If both experts accept their respective contracts, they deliver odds $f_1$ and $f_2$ and when the final state $s$ is observed, expert 1 receives payoff $C_1(f_1, f_2, s)$ (or gives this ammount when the payoff is negative).

When expert 1 is informed, he maximizes his expected utility conditional on the other experts forecast. We say that expert 1 *accepts the contract* if for every $f_1, f_2 \in \Delta(S)$ we have $E^{f_1}\{C_1(f_1, f_2, \cdot)\} > 0$. That is, when revealing the truth gives him a positive payoff regardless of the other expert's forecast. Moreover, we say he *honestly reveals his beliefs* when for all $f_2 \in \Delta(S)$ and $f' \neq f_1 \in \Delta(S)$

$$E^{f_1}\{C_1(f_1, f_2, \cdot\,)\} > E^{f_1}\{C_1(f', f_2, \cdot\,)\}.$$

[13] In the case where only one expert accepts the contract the tester needs a mechanism that elicits the truth, he can replace the second experts forecast with a prior like in Sandroni (2014) and achieve this result, this is not the main focus of our results.

This property ensures that informed experts don't have incentives to misrepresent their beliefs and can be achieved using proper scoring rules.

We say that expert 1 is uninformed when he doesn't know the true data generating process and he only considers possible true theories those in the subset $\Theta_1 \subset \Delta(S)$. We suppose $\Theta_1$ is a closed set and it contains at least 2 points. If uninformed, expert 1 evaluates his prospects using the minmax criteria as described by Gilboa and Schmeidler (1989). Considering both experts may announce their odds using random generators of theories $\xi_1$ and $\xi_2 \in \Delta(\Delta(S))$ this can be stated as saying that uninformed expert 1 only accepts contract $C_1$ when:

$$\max_{\xi_1 \in \Delta(\Delta(S))} \min_{\substack{f \in \Theta_1 \\ \xi_2 \in \Delta(\Delta(S))}} \int_{\Delta(S)} \int_{\Delta(S)} E^f C^1(f', f^*, \cdot) \xi_2(f^*) \xi_1(f') > 0. \tag{10}$$

If expert 1 is uninformed, we say that he rejects the contract $C_1$ when there's no $\xi_1 \in \Delta(\Delta(S))$ that satisfies (11).

**Proposition 9.** In a multiple expert setting assume that $\Theta_1$ contains at least two points. There's a contract $C_1$ such that expert 1 if informed accepts it and reveals his knowledge and if uninformed rejects it.

The intuition of the proof is very simple. We build a contract that gives a payoff proportional to the difference of a proper scoring rule plus a premium of $\varepsilon$. True experts can be assured to get paid at least $\varepsilon$ since proper scoring rules are maximized only when saying the truth. False experts are victims of the same reasoning, in case the other expert is informed, their scoring rule will be dominated by that of the other expert and cause them loses that cannot be compensated by the premium $\varepsilon$. In order to guarantee a positive payment they would need to randomize in a way that their forecasts are always close to the true odds and this is impossible.

We can symmetrically define a contract for expert 2. This achieves perfect screening as informed experts would always accept the contract and reveal the truth, and there's always a contract that an uninformed expert would reject. There's a caveat to this result, because the screening contract depends on the diameter of the set $\Theta_1$. So this solution does not give a single contract that screens all uninformed experts, but rather for each set of experts a contract that screens informed from uninformed experts exists. The proof is as follows:

*Proof.* Consider a continuous strictly proper scoring rule $B : \Delta(S) \times S \to \mathbb{R}$ (e.g. Brier (1950)), that is to say that for all $f' \neq f \in \Delta(S)$:

$$E^f(B(\mathbf{f}, s)) > E^f(B(\mathbf{f'}, s)).$$

Given $\varepsilon > 0$ we define a contract $C_1$ as $C_1(f_1, f_2, s) = B(f_1, s) - B(f_2, s) + \varepsilon$

Because $B$ is a strictly proper scoring rule, we have that $\mathbb{E}^f C_1(f, f_2, \cdot) \geq \varepsilon$, so informed experts always accept the contract. Moreover, for the same reason we have that, for $f' \neq f_1$:

$$E^{f_1} C_1(f_1, f_2, s) = E^{f_1} B(f_1, s) - E^{f_1} B(f_2, s) + \varepsilon > E^{f_1} B(f', s) - E^{f_1} B(f_2, s) + \varepsilon = E^{f_1} C_1(f', f_2, s),$$

so informed experts honestly reveal the truth.

If expert 1 is uninformed begin by noting that his maxmin payoff is inferior to the one obtained if the other expert forecasts the true probability (which he would if he's informed). The payoff in such case is:

$$\max_{\xi_1 \in \Delta(\Delta(S))} \min_{f \in \Theta_1} \int_{\Delta(S)} E^f C^1(f', f, \cdot) \xi_1(f')$$

The remainder of the proof consists of proving there is a value of $\varepsilon$ small enough that the above expression is negative. Note that the function $E^f \left( B(f, s) - B(f', s) \right)$ is continuous on $f'$. This follows because the expected value is a weighted average of $B$, a continuous function of $f'$).

Now, take two distinct points $f_x, f_y \in \Theta_1$. Note that the term $E^f \left( B(f', s) - B(f, s) \right)$ is always non-negative, since $B$ is a proper scoring rule. Consider the distance (euclidean distance in $[0, 1]^R$, with $R = |S|$) between $f_x$ and $f_y$ and call it $d$. Now consider the open ball $b_x := \{f \in \Theta || f - f_x|| < \frac{d}{2}\}$.

Since $b_x$ is open, it is complement $b_x^{\mathbf{c}}$ is a closed subset of the compact $[0, 1]^R$, hence it is compact. It follows that the continuous non-negative function $E^{f_x} \left( B(f', s) - B(f_x, s) \right)$ attains a minimum in $b_{\frac{d}{2}}(f_x)^{\mathbf{c}}$, namely $\delta_x$. Since $B$ is strictly proper it follows that $\delta_x > 0$.

We proceed by contradiction. Suppose there's a random generator of theories

$\bar{\xi} \in \Delta(\Theta)$ with which uninformed expert 1 accepts the contract. it is easy to see that $\bar{\xi}(b_x) > \frac{1}{2}$, for this note that if $\bar{\xi}(b_x^{\mathbf{c}}) \geq \frac{1}{2}$, then the expert's payoff (when the other expert forecasts the truth) is less or equal than:

$$\begin{aligned}
\min_{f\in\Theta} \int_{\Theta} E^f C^1(f', f, \cdot)\bar{\xi}(f') &\leq \int_{\Theta} E^{f_x} C^1(f', f_x, \cdot)\bar{\xi}(f') \\
&= \int_{b_x^{\mathbf{c}}} E^{f_x} C^1(f', f_x, \cdot)\bar{\xi}(f') + \int_{b_x} E^{f_x} C^1(f', f_x, \cdot)\bar{\xi}(f') \\
&< \bar{\xi}(b_x^{\mathbf{c}})(\varepsilon - \delta_x) + \frac{\varepsilon}{2},
\end{aligned}$$

the last inequality coming from the definition of $b_x$ and from the fact that $C$ is bounded from above by $\varepsilon$.

For $\varepsilon < \frac{\delta_x}{2}$ the last term is negative, resulting in a contradiction since with $\bar{\xi}$ the expert supposedly accepts the contract, so we must have $\bar{\xi}(b_x) > \frac{1}{2}$. In an analogous manner it can be shown that if $b_y = \{f \in \Theta | ||f - f_y|| < \frac{d}{2}\}$, then there's $\delta_y > 0$ such that $\delta_y = \min_{b_y^{\mathbf{c}}} E^{f_y}\left(B(f', s) - B(f^y, s)\right)$ and for $\varepsilon < \frac{\delta_y}{2}$ we must have $\bar{\xi}(b_y) > \frac{1}{2}$. Thus taking $\varepsilon < \min\left(\frac{\delta_x}{2}, \frac{\delta_y}{2}\right)$ yields a contradiction because $b_x$ and $b_y$ are disjoint and then $\bar{\xi}$ cannot exist.

□

The structure of the contract is very similar to that given by Sandroni (2014), and it also only requires a single data point for it to work. The idea of screening experts based on one forecast overperforming the other bears resemblance to the test in Al-Najjar and Weinstein (2008), but in this case it is not necessary for the tester to know for sure that a true expert is present. In fact this test would still be rejected by two uninformed experts if the assumption that they are uncertain about the other's type is made.

The assumption that they are uncertain about the other expert's type cannot be relaxed, since if the two experts are uninformed and know about each other's type they could ensure themselves a positive payoff by playing identically (each receives a payment of $\varepsilon$). However, in the proof of proposition 9 it is only necessary for each expert to consider possible that the other expert forecasts the true odds. So it is only necessary to assume that uninformed experts are not certain that the other expert cannot forecast the truth, a very reasonable assumption.

The previous result comes from the fact that an uninformed expert cannot always be close to the true distribution even when randomizing. It is worth noting that when the uninformed expert considers a set of theories very close to the truth as plausible, then the value of $\varepsilon$ needs to be sufficiently small. This implies that the principal has no way to screen all uninformed experts, rather she can devise contracts that are only accepted by experts that are sufficiently informed in the sense that their predictions are at a certain distance from the truth. This type of contracts can be exploited to screen experts that are better informed than others as we explain below.

## 3.4 Contracts that choose the more informed expert

We now consider a setting involving partially informed experts. Partially informed experts are defined following Olszewski and Sandroni (2007), as experts that almost perfectly know the true odds of the states. Formally expert $i$ is partially informed if he's uninformed and his set of plausible theories $\Theta_i$ takes the form $B_\delta(f^*) = \{f \in \Delta(S) : ||f - f^*|| \leq \delta\}$ (closed ball centered at $f^*$), where $f^*$ are the true odds. We say that an expert is better informed than other when their sets of plausible theories are $B_\delta(f)$ and $B_{\delta'}(g)$ and $\delta < \delta'$ for some pair of theories $f$ and $g$. Notice that we do not require their sets of plausible theories to be centered at the true theory, but naturally it is required that they do contain it.

In this setting we make explicit use of the Brier score and some of it is properties.

**Definition 12.** The Brier Score is a scoring rule $B : \Delta(S) \times S \to \mathbb{R}$, defined as $B(f,s) = 2f(s) - \sum_{s' \in S}(f(s'))^2 - 1$

**Lemma 4.** The Brier Score satisfies the following equation:

$$E^f\{B(g,\cdot\ )\} = ||f||_2^2 - ||f-g||_2^2 - 1, \tag{11}$$

where $||\cdot||_2^2$ denotes the $\mathcal{L}_2$ norm squared.

**Lemma 5.** The Brier Score is a continuous function and it is a proper scoring rule. That is to say that for every $g \in \Delta(S)$ such that $g \neq f$ we have:

$$E^f\{B(f,\cdot\ )\} = E^f\{B(g,\cdot\ )\}, \tag{12}$$

Proofs of these two lemmas are presented in the appendix. Next we prove a result that shows that when an expert is even slightly better informed than another one, there's a contract that separates them perfectly.

**Proposition 10.** In a setting with two partially informed experts, where one is better informed than the other there's a contract that is accepted by the better informed expert and rejected by the other.

The assumption that there's one expert that has perfect knowledge and another that is completely ignorant is both extreme and unrealistic. Proposition 10 shows us that even in the case when one expert is slightly better informed, there's a contract that achieves perfect screening. Proposition 9 might be regarded as a degenerate case of proposition 10. The proof is as follows:

*Proof.* Let $\Theta_1 = B_{\varepsilon_1}(f_x)$ and $\Theta_2 = B_{\varepsilon_2}(f_y)$ for a pair of theories $f_x$ and $f_y$, and without loss of generality assume $\varepsilon_2 > \varepsilon_1 > 0$. $B$ will be the Brier Score over the set $S$. Let $\gamma$ be an intermediate value between $\varepsilon_1$ and $\varepsilon_2$, that is $\varepsilon_2 > \gamma > \varepsilon_1$ Consider the following contracts for experts 1 and 2 respectively:

$$C_1(f_1, f_2, s) = B(f_1, s) - B(f_2, s) + \gamma^2,$$
$$C_2(f_2, f_1, s) = B(f_2, s) - B(f_1, s) + \gamma^2.$$

Applying lemma 4 we know that:

$$E^f\{C_1(f_1, f_2, \cdot\ )\} = ||f_2 - f||_2^2 - ||f_1 - f||_2^2 + \gamma^2, \tag{13}$$
$$E^f\{C_2(f_2, f_1, \cdot\ )\} = ||f_1 - f||_2^2 - ||f_2 - f||_2^2 + \gamma^2. \tag{14}$$

Since each expert faces uncertainty they evaluate their prospects according to Gilboa and Schmeidler's (1989) minmax criteria, so provided that experts may use random generators of theories $\xi_1, \xi_2 \in \Delta(\Delta(S))$ they solve:

$$\max_{\xi_i \in \Delta(\Delta(S))} \min_{\substack{f \in \Theta_i \\ \xi_j \in \Delta(\Delta(S))}} \int_{\Delta(S)} \int_{\Delta(S)} E^f C_i(f', f^*, \cdot) d\xi_i(f^*) d\xi_j(f'), \tag{15}$$

for $i \neq j$.

Now, it is trivial to see that the worst-case scenario for expert $i$ is when $\xi_j(\{f\}) = 1$, since it minimizes the first term in (3) and (4). So expert $i$ needs to solve:

$$\begin{aligned}
\max_{\xi_i \in \Delta(\Delta(S))} \min_{\substack{f \in \Theta_i \\ \xi_j \in \Delta(\Delta(S))}} & \int_{\Delta(S)} \int_{\Delta(S)} E^f C_i(f', f^*, \cdot) d\xi_i(f^*) d\xi_j(f') \\
&= \max_{\xi_i \in \Delta(\Delta(S))} \min_{f \in \Theta_i} \int_{\Delta(S)} E^f C_i(f', f, \cdot) d\xi_i(f') \\
&= \max_{\xi_i \in \Delta(\Delta(S))} \min_{f \in \Theta_i} \int_{\Delta(S)} \left(\gamma^2 - ||f - f'||_2^2\right) d\xi_i(f') \\
&= \max_{\xi_i \in \Delta(\Delta(S))} \min_{f \in \Theta_i} \gamma^2 - \int_{\Delta(S)} ||f - f'||_2^2 d\xi_i(f').
\end{aligned}$$

It is simple to see that expert 1 accepts his contract, because he can get a positive payoff by making $\xi_1(\{f_x\}) = 1$. Or formally:

$$\begin{aligned}
\max_{\xi_1 \in \Delta(\Delta(S))} \min_{f \in \Theta_1} \gamma^2 - \int_{\Delta(S)} ||f - f'||_2^2 d\xi_1(f') &\geq \min_{f \in \Theta_1} \gamma^2 - ||f - f_x||_2^2 \\
&\geq \min_{f \in \Theta_1} \gamma^2 - \varepsilon_1^2 > 0,
\end{aligned}$$

the second inequality coming from the fact that $\Theta_1 = B_{\varepsilon_1}(f_x)$.

Seeing that expert 2 rejects his contract is less obvious, since he could randomize over $\Theta_2$ to try to on average be closer than $\gamma$ of the true odds. We must show that any randomization has a point that is on average further than $\gamma$, similar to what happens when expert 2 uses a pure strategy.

We proceed by contradiction. Let $\bar{\xi_2} \in \Delta(\Theta_2)$ be a strategy by expert 2 for which he accepts the contract. Formally this is:

$$\min_{f \in \Theta_2} \gamma^2 - \int_{\Delta(S)} ||f - f'||_2^2 d\bar{\xi_2}(f') > 0. \tag{16}$$

Let $\Phi : \Delta(S) \to \mathbb{R}^{>0}$ be the function such that $\Phi(g) = ||g||_2^2$. This function is convex and it is defined over a convex vector space, so applying Jensen's inequality we get:

$$\left\| f - \int_{\Delta(S)} f' d\bar{\xi_2}(f') \right\|_2^2 = \left\| \int_{\Delta(S)} f - f' d\bar{\xi_2}(f') \right\|_2^2 \leq \int_{\Delta(S)} \|f - f'\|_2^2 d\bar{\xi_2}. \tag{17}$$

So if we make $\bar{f} = \int_{\Delta(S)} f' d\bar{\xi_2}(f')$, we know that $\bar{f} \in \Delta(S)$ since it is a convex set. And so we get:

$$\begin{aligned}
&\min_{f\in\Theta_2} \gamma^2 - \int_{\Delta(S)} ||f - f'||_2^2 d\bar{\xi_2}(f') \\
&\qquad \le \min_{f\in\Theta_2} \gamma^2 - \left\|f - \bar{f}\right\|_2^2 \\
&\qquad = \gamma^2 - \max_{f\in\Theta_2} \left\|f - \bar{f}\right\|_2^2 \\
&\qquad \le \gamma^2 - \varepsilon_2^2 < 0.
\end{aligned}$$

Which is a contradiction with (17), so $\bar{\xi_2}$ does not exist. The fact that $\max_{f\in\Theta_2} \left\|f - \bar{f}\right\|_2 \ge \varepsilon_2$ was used, and it follows from noting that there are two points in $\Theta_2$ at a distance of $2 * \varepsilon_2$ and applying triangular inequality. □

## 3.5 Discussion

The results in proposition 9 need to be compared with other results in the literature. The existence of a contract that perfectly screens experts is a result in the opposite direction to that of Olszewski and Sandroni (2009a). A contract setting allows the principal to take profit of the strategical uncertainty faced by uninformed experts. Nonetheless, it is clear that coordination between uninformed experts remains a problem [14], but in this case the possibility of facing an informed expert is enough to reject the contract.

This contract is also similar and inspired in the positive results obtained by Al-Najjar and Weinstein (2008) and Dekel and Feinberg (2008), in the sense that the presence of an informed expert in the game (in this case the possibility of this) is what makes manipulation impossible. However there are two main improvements over these results. First, tests like the one in Al-Najjar and Weinstein (2008) require infinite periods to screen informed and uninformed expert. This test achieves screening on a single data point. Second, in this contract the principal does not need to know that there's an informed expert in the game to achieve screening. it is enough for uninformed experts to be uncertain about the other expert's type, a weaker and more realistic assumption.

[14] Two uninformed experts can get positive payoffs if they forecast the same odds in a contract like the one in the proof of proposition 9

This result is similar and greatly inspired by that of Proposition 6, with the difference that the presence of a second expert allows to dispense of the assumption that the principal has a prior which is conceivable for uninformed experts. The fear of being over-performed by another expert is enough to disincentivize strategic forecasting and avoiding harmful odds.

The difference in results when considering models with multiple experts sheds some light on the subtle circumstances that allow for manipulation results in the single expert setting. As noted by Al-Najjar and Weinstein (2008), in single expert testing the space of strategies is too small compared to that of the expert. Requiring that tests pass the truth makes any mixed strategy by Nature payoff equivalent to a pure strategy, eliminating any meaningful sense in which Nature can randomize. The presence of a potentially informed expert gives Nature a larger strategy space in the sense that the uninformed expert no longer has a good response to any mixed strategy by Nature.

The results of proposition 9 and 10 show that, effectively, the presence of another expert reduces the game to one where an uninformed expert would need to be near the true distribution to get a good payoff. An analogy can be made to the scientific method, it is much easier to prove that a theory performs better than other than to prove that a certain theory is wrong or right. Precisely this is what we observe in different fields of science. Theories frequently show inconsistencies with data but are maintained because they are the best proposed theory so far.[15] With the simple mechanism we propose we can screen true theories from false ones by means of comparing their performance, a simple idea that is inaccessible when evaluating in isolation.

As a conclusion we point out that the impossibility results in single expert testing that can be overcome in the multiple expert setting show how insufficient data can be to validate theories. The different results presented here show that only comparing with the theory of a better informed expert can we reveal the quality of a theory.

[15]Common examples are the Newtonian theory of gravity, general relativity, and expected utility theory.

# 4 Appendix

## 4.1 Example of manipulation of R-test

This test is known as R-test y formally it states the following; if $f^P_{t-1}(s)$ is the forecasted probability for period $t$ based on history $s$, and $I_t(s)$ is the indicator function of event $\{1\}$ in period $t$, then Alice rejects Bob's theory on path $s$ unless:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}(f^P_{t-1}(s) - I_t(s)) = 0$$

Figure 1: The criterion to accept Bob's theory

Can Bob, knowing he's going to be tested in this manner, pass the test if completely ignorant of the real probabilities? The answer is affirmative. Bob can adjust his predictions depending on how calibrated he's been in the past history. Bob can guarantee himself to always approach the limit value in the rejection criterion. Let $R(f, s_t) = \sum_{i=1}^{t}(f^P_{i-1}(s) - I_t(s))$ then the following strategy works for Bob:

$$\begin{aligned} f(s_t) &= 1 \quad \text{si } R(f, s_{t-1}) < 0 \\ f(s_t) &= \frac{1}{2} \quad \text{si } R(f, s_{t-1}) = 0 \\ f(s_t) &= 0 \quad \text{si } R(f, s_{t-1}) > 0 \end{aligned}$$

With this strategy the term $R$ is always approaching 0, and it is trivial to see that the limit converges to 0. Thus showing that although Bob produces useless forecasts, he can manipulate tests.

## 4.2 Proof of lemmas 4 and 5

**Proof of lemma 4**

$$
\begin{aligned}
&E^f\{B(g,\cdot\ )\} \\
&\qquad = -\sum_{s\in S} f(s)\left(1 - 2g(s) + \sum_{s'\in S}(g(s'))^2\right) \\
&\qquad = \sum_{s'\in S}(g(s'))^2 + \sum_{s\in S} 2f(s)g(s) - 1 \\
&\qquad = ||f||_2^2 - ||f-g||_2^2 - 1
\end{aligned}
$$

**Proof of lemma 5**

*Proof.* This follows immediately from lemma 4 and the fact that $||\cdot||_2$ is a norm. □

## 4.3 Example of challenges in manipulating comparative tests

**Example 1** (Olszewski and Sandroni (2009a)). Given a positive number $k$, we say that theory $f$ $k-$outperforms theory $g$ at history $h^t$ if

$$\frac{P^f(h^t)}{P^g(h^t)} \geq k,$$

in other words, history $h^t$ is at least $k$ times more likely according to theory $f$ than according to theory $g$.

A useful notion of 'closeness' between theories is discussed in (Al$-$najjar, 2008). Given $\eta > 0$ and a natural number $r$. For a given history $h_t$, let $h_t|s$ be the $s$ history whose outcomes coincide with the first $s-1$ outcomes of history $h_t$. A theory $f$ is $(\eta, r)$ similar to theory $g$ if at history $h_t$ when:

$$|f(h_t|s) - g(h_t|s)| < \eta,$$

for all $s = 1, \ldots, t$ except for at most $r$ of them.

Given any theory $f$, the tester might want to compare it to an alternate theory which is different in a way. One such theory $\bar{f}$ can be constructed taken a $\gamma \in (0, .5]$ defined as:

$$\bar{f}(h^t) = \begin{cases} f(h^t) + \gamma \text{ if } f(h^t) < 0.5 \\ f(h^t) - \gamma \text{ if } f(h^t) \geq 0.5 \end{cases}$$

So the forecasts of a theory differ by $\gamma$ of those of its alternate theory $\bar{f}$.

We define a test $\bar{T}$ as follows; fix numbers $k > 1$, $\eta > 0$, $\gamma \in (0, .5]$ and positive natural numbers $r$ and $m$. The rejection sets of $\bar{C}$ consist only of $m-$histories. Theory $f^i$ of expert $i$ is rejected at history $h_m$ if:

1. The theory $f^i$ does not $k-$ outperform the alternate theory $\bar{f}^i$ at $h_m$;

   or if

2. The theory $f^i$ is not $(\eta, r)$ similar to theory $f^j$ of expert $j \neq i$ and theory $f^j$ $1-$outperforms theory $f^i$ at $h_m$.

Condition 1 by itself is the same as a likelihood test which has been studied in (Sandroni,2008). Condition 2 is an adaptation of the test studied in Al$-$Najjar and Weinstein (2008); which can detect the true expert when predictions are not similar and the tester knows of the presence of a true expert. In this setting, this is not case, and the tester must deal with the possibility of both experts being false.

By proposition 2 in (Sandroni, 2008) and proposition 1 in (Al$-$najjar, 2008) we know that for any $k, \gamma$ and $\varepsilon > 0$ there exist naturals $m$ and $r$ such that test $\bar{C}$ passes the truth with probability $1-\varepsilon$. it is also easy to see that this test is future independent, as it relies on predictions made up to period $m$. So by theorem 3, we know that such a test can be ignorantly passed with probability $1 - \varepsilon - \delta$ for arbitrarily small $\delta$.

However, note that such manipulation of the test is not an easy task. If we fix an expert, we see that no matter what theory $f$ he chooses, the alternate theory $\bar{f}$ outperforms it in several datasets. So if this expert announces any theory deterministically, he will be rejected at some $m-$histories. This implies that both experts need to randomize with very specific odds to avoid rejection by condition 1.

In addition, on any dataset at least one expert fails the test if the theories are not similar. Thus, if the test is going to be ignorantly passed, both experts theories need to be similar with high probability. However, with a large $m$ (compared to r) it would be difficult to announce similar theories in most cases since they have to randomize independently. So there's a conflict between the two conditions, condition 1 requires that they randomize their theories carefully and condition 2 requires their theories to be similar (which is hard when independently randomizing); however we are guaranteed by Proposition 8 that such randomizations exist, although the construction of such randomization cannot be inferred from the proof.

# References


[1] Nabil I Al-Najjar and Jonathan Weinstein. Comparative testing of experts. *Econometrica*, pages 541–559, 2008.

[2] Glenn W Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

[3] Andrés Carvajal. Statistical calibration: A simplification of fosters proof. *Mathematical Social Sciences*, 58(2):272–277, 2009.

[4] A Philip Dawid. The well-calibrated bayesian. *Journal of the American Statistical Association*, 77(379):605–610, 1982.

[5] Eddie Dekel and Yossi Feinberg. Non-bayesian testing of a stochastic prediction. *The Review of Economic Studies*, 73(4):893–906, 2006.

[6] Ky Fan. Minimax theorems. *Proceedings of the National Academy of Sciences*, 39 (1):42–47, 1953.

[7] Yossi Feinberg and Colin Stewart. Testing multiple forecasters. 2007.

[8] Dean P Foster and Rakesh V Vohra. Asymptotic calibration. *Biometrika*, 85(2): 379–390, 1998.

[9] Itzhak Gilboa and David Schmeidler. Maxmin expected utility with non-unique prior. *Journal of mathematical economics*, 18(2):141–153, 1989.

[10] Irving L Glicksberg. A further generalization of the kakutani fixed point theorem, with application to nash equilibrium points. *Proceedings of the American Mathematical Society*, 3(1):170–174, 1952.

[11] Amy Greenwald and Amir Jafari. Blackwells approachability theorem: A generalization in a special case. 2006.

[12] Wojciech Olszewski and Marcin Peski. The principal-agent approach to testing experts. *American economic Journal: microeconomics*, 3(2):89–113, 2011.

[13] Wojciech Olszewski and Alvaro Sandroni. Contracts and uncertainty. *Theoretical Economics*, 2(1):1–13, 2007.

[14] Wojciech Olszewski and Alvaro Sandroni. Strategic manipulation of empirical tests. 2008.

[15] Wojciech Olszewski and Alvaro Sandroni. Manipulability of comparative tests. *Proceedings of the National Academy of Sciences*, 106(13):5029–5034, 2009.

[16] Wojciech Olszewski and Alvaro Sandroni. A nonmanipulable test. *The Annals of Statistics*, pages 1013–1039, 2009.

[17] Wojciech Olszewski et al. Calibration and expert testing. *Handbook of Game Theory with Economic Applications*, 4:949–984, 2015.

[18] Alvaro Sandroni. The reproducible properties of correct forecasts. *International Journal of Game Theory*, 32(1):151–159, 2003.

[19] Alvaro Sandroni. At least do no harm: The use of scarce data. *American Economic Journal: Microeconomics*, 6(1):1–4, 2014.