Strategic Manipulation of Empirical Tests*

Wojciech Olszewski† and Alvaro Sandroni‡

May 10, 2006

Abstract

Theories can be produced by individuals seeking a good reputation of knowledge. Hence, a significant question is how to test theories anticipating that they might have been produced by (potentially uninformed) experts who prefer their theories not to be rejected.

If a theory that predicts exactly like the data generating process is not rejected with high probability then the test is said to not reject the truth. On the other hand, if a false expert, with no knowledge over the data generating process, can strategically select theories that will not be rejected then the test can be ignorantly passed. These tests have limited use because they cannot feasibly dismiss completely uninformed experts.

Many tests proposed in the literature (e.g., calibration tests) can be ignorantly passed. Dekel and Feinberg (2006) introduced a class of tests that seemingly have some power of dismissing uninformed experts. We show that some tests from their class can also be ignorantly passed. One of those tests, however, does not reject the truth and cannot be ignorantly passed. Thus, this empirical test can dismiss false experts.

*We thank Eddie Dekel, Yossi Feinberg, Nabil Al-Najjar, Sandeep Baliga, Roger Myerson and Rakesh Vohra for useful comments. All errors are ours.

†Department of Economics, Northwestern University, 2001 Sheridan Road Evanston IL 60208

‡Department of Economics, University of Pennsylvania, 3718 Locust Walk, Phildadelphia PA 19104 and Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road Evanston IL 60208

We also show that a false reputation of knowledge can be strategically sustained for an arbitrary, but given, number of periods, no matted which test is used (provided that it does not reject the truth). However, false experts can be discredited, even with bounded data sets, if the domain of permissible theories is mildly restricted.

1. Introduction

Economists have long realized that the production and the transmission of knowledge play a central role in economic activity (see Hayek (1945)). Knowledge, however, is often structured as a theory which must be tested. A theory can be rejected if it makes a deterministic prediction that is not observed in the data. Still, in economics and several other disciplines, theories regularly make probabilistic forecasts that attach strictly positive probability to all outcomes. This leads to the basic question of how to test probabilistic theories. If a blunt contradiction between the theory and the data is impossible then the standard procedure is to employ large data sets so that any theory must attribute small probability to some events which, if observed, induce a rejection of the theory. A matter of interest is precisely which low probability events should be regarded as sufficiently incompatible with the theory to validate its rejection.

Assume that the problem at hand requires an understanding of a stochastic process which generates an outcome that can either be 0 or 1. Before any data is observed, a potential expert named Bob delivers a theory, defined as a mapping that takes as an input any finite string of outcomes and returns as an output a probability of 1. There is a well-known correspondence between theories and probability measures P over the space of infinite sequences of outcomes. So, henceforth, a theory is a probability measure P. Bob may be an informed expert who truthfully reveals the data generating process. However, Bob may also be a false expert who knows nothing about the data generating process.

A tester named Alice tests Bob's theory P by selecting an event A_P (i.e., a set of outcome sequences). Alice regards the event A_P as consistent with the theory P and its complement A_P^c as inconsistent with it. Assume that the acceptance set A_P has high probability according to P, i.e.,

$$P(A_P) \geq 1 - \varepsilon. \tag{1.1}$$

Then, if Bob's theory coincides with the data generating process, it will not be rejected with probability $1 - \varepsilon$. When equation (1.1) is satisfied we say that Alice's test does not reject the truth with probability $1 - \varepsilon$.

A vast effort has been dedicated to supply results that take the form of equation (1.1). The significance of these results (such as the law of large numbers, the law of iterated logarithmic, the central limit theorem) is that they relate the unobservable concept of a theory P with a potentially observable event A_P. So, each of these findings can be used to define an empirical test that does not reject the truth.

An essential feature of a test is the possibility of theory rejection. If it is certain that a theory will not be rejected then the purpose of the test is far from clear. As long as A_P^c is a non-empty set, theory rejection is (seemingly) viable because outcome sequences in A_P^c may be realized. However, even if the rejection set A_P^c is non-empty Bob might still be able to strategically produce theories in a way that essentially removes the possibility of rejection.

This might be accomplished without any knowledge of the data generating process. Assume that Bob, before any data is observed, uses a random device ζ to select his theory P. At first, Alice cannot tell whether the theory she received was revealed truthfully by an informed expert or was selected at random by an ignorant expert. This must be determined by the data. However, assume that for *any* sequence of outcomes Bob's theory P will not be rejected with arbitrarily high probability, according to Bob's randomization device ζ. It follows that, no matter which data is realized, Alice will not reject Bob's theory (unless Bob had an unlucky draw from ζ which is, by definition, near impossible). If such a device ζ can be constructed, we say that the test can be ignorantly passed. On the other hand, if for every random generator of theories ζ there exists at least one sequence of outcomes such that, if realized, leads to theory rejection (with high probability according to ζ) then we say that the test cannot be manipulated. A test that does not reject the truth with high probability and cannot be manipulated is called an effective test.

As aforementioned, a test that is not effective has limited purpose if Bob produces theories strategically. Alice has no reason to run a test that delivers the same verdict to both polar cases of an expert who knows the data generating process and a false expert who knows nothing about it. However, a test that is effective has hitherto not been proven to exist.[1] The inexistence of an effective test would impose a conceptual limit on the capability of empirical analysis. Theorems in probability can be demonstrated and subsequently used to construct statistical tests that do not reject the truth, but the value of any test is doubtful if it can be ignorantly passed. Hence, a fundamental question is whether an effective test exists.

The calibration test requires the empirical frequency of 1 to be close to p in the periods that 1 was forecasted with probability close to p. Foster and Vohra (1998) show that the calibration test can be ignorantly passed with probability one. So, it is possible to produce forecasts that in the future will prove to be calibrated, no matter which data sequence is eventually realized.

The Foster and Vohra (1998) result has been extended and several other tests have been suggested (see, for example, Lehrer (2001)).[2] Still, these tests can be ignorantly passed with probability one. However, the literature has not yet explored all possible tests.

Natural candidates for effective tests are associated with small acceptance sets. By definition, the acceptance set of a category test is a first category set, i.e., a countable union of nowhere dense sets. Such sets are considered small in topology. A

[1] The only exception is a test suggested by Dekel and Feinberg (2006), but this test is based on the continuum hypothesis and, hence, it is not proven to exist with standard axioms.

[2] We also refer the reader to Cahn (2005), Foster and Young (2003), Fudenberg and Levine (1995, 1999a, 1999b), Hart (2005), Hart and Mas-Colell (2000, 2001), Lehrer and Solan (2003), Rustichini (1999), Kalai, Lehrer and Smorodinsky (1999), Sandroni, Smorodinsky and Vohra (2003), and Vovk and Shafer (2004) for related work.

classic result shows the existence of category tests that do not reject the truth with probability one. Dekel and Feinberg (2006) show that any theory fails with probability one any category test, unless the data generating process belongs to a first category set (that includes the announced theory). So, Bob's prospects of passing any category test may seem negligible, but they are not. We show that some category tests can be ignorantly passed.

A theory is said falsifiable if, for every finite history, there is a finite continuation history that the theory assigns probability zero to it. The falsifiability test rejects out of hand (i.e., on all paths) any non-falsifiable theory. A falsifiable theory is rejected if and only if it assigns zero probability to the observed (finite) data. We show that the falsifiability test is a category test. Moreover, we show that the falsifiability test can be ignorantly passed with arbitrarily high probability.

Bob, if uninformed, does not know the likelihood that any given falsifiable theory will be falsified. In addition, no matter which theory he announces, it will be rejected with probability one for all actual data generating processes, except for a few of them (i.e., a first category set). However, Bob can generate falsifiable theories at random so that, for any data, the realized theory will not be falsified, with arbitrarily high probability. By themselves, the falsifiability criteria are powerless to turn out an effective test.

Empirical tests can be classified into two kinds: acceptance tests and rejection tests. An acceptance test accepts a theory in finite time when the data is judged consistent enough with it. Rejection may occur at infinity. A rejection test rejects a theory in finite time when the data is judged inconsistent enough with it. Acceptance may occur at infinity. A test that is both an acceptance and rejection test must stop at some finite time (that depends upon the announced theory) and either reject or accept the theory.

We show that any acceptance test that does not reject the truth with high probability can be ignorantly passed with high probability. Hence, consider an arbitrary definition of which set of finite data is consistent with any given theory. Assume that, with high probability, the test will not regard a theory that predicts exactly like the data generating process as inconsistent with the data. Then, with no knowledge over the data, it is possible to produce theories that will prove to be consistent with the data.

In contrast, we construct a rejection test, called the global category test, that does not reject the truth with arbitrarily high probability and that cannot be manipulated. Hence, an effective empirical test does exist. This result reduces the conceptual bound that would otherwise (if an effective test did not exist) be imposed on the purpose of testing probabilistic theories. For any given theory, we can identify a set of finite data, deemed as inconsistent with it, which allows for the possibility of rejecting false theories, even if they are strategically produced.

The global category test discredits uninformed experts in a strong sense. Given any random generator of theories ζ, let the revelation set consists of the paths on which ζ fails the test with probability one. We show that for every ζ, the revelation set of the global category test is not only non-empty (which makes rejection of false experts feasible), but also large enough to have full measure for all probability measures, save a first category set of them.

The critical difference between the global category test and the tests proposed in the literature (e.g., calibration tests) is that the latter cannot dismiss false experts, but the former can. This distinction is enhanced by the finding that an uniformed expert, if tested by the global category test, experiences inevitable failure on a topologically large set of paths. So, the dismissal of an uninformed expert is both feasible and plausible.

Our results imply an asymmetry between acceptance and rejection tests: No acceptance test is effective, but some rejection tests are effective. This asymmetry merits some emphasis (although it departs from our main subject) because it gives formal support to a guiding principle: theories must be accepted until revealed to be in conflict with the data rather than rejected until shown to be in agreement with it.

We now turn to the question of how long it takes to discredit false experts. We say that rejection can be delayed for m periods if theories can be strategically generated at random so that, for any sequence of outcomes, the realized theory will not be rejected before period m with high probability, according to the randomization device. We show that for any period m, and for any test that does not reject the truth with high probability, rejection can be delayed for m periods. Thus, Bob may not be able to eternally sustain a false reputation of being knowledgeable, but Bob can maintain a false reputation within an arbitrarily long time horizon. It follows that, with bounded data sets, it might be unfeasible for Alice to obtain any significant information on whether Bob is a knowledgeable expert.

The fact that false experts can delay rejection poses a severe difficulty for the empirical testing of probabilistic theories. This difficulty seems unavoidable unless we relax the assumption that the test does not reject the truth. Thus, we allow tests to reject some theories out of hand (i.e., on all data sets and, therefore, independently of the data). This is not a desiderata, but it is a necessary recourse.

We construct an empirical test that rejects out hand the theories that frequently forecast 1 and 0 with probability near 0.5. The test requires the empirical frequency of each outcome to be high in the periods that they were forecasted with probability appreciably greater than 0.5. This test is related, but not identical, to tests used for meteorological forecasts and also to calibration tests. We show that this test does not reject, with high probability, any allowed theory.

Moreover, Bob, if uninformed, cannot assure himself a long rejection delay of his theories. Thus, a rejection of strategically produced theories is feasible with bounded data sets if we impose mild restrictions on the class of permitted theories.

The paper is organized as follows: In section 2 we introduce our main concepts, show some examples of empirical tests, and demonstrate that acceptance tests are not effective. In section 3, we show that the falsifiability test can be ignorantly passed. In section 4, the global category test is presented and its properties are demonstrated. In section 5, we show that if the truth is not rejected then rejection can be arbitrarily delayed. In section 6, we construct a test such that a false expert cannot ensure a long reputation of knowledge. Section 7 concludes the paper. Proofs are in the appendix.

2. Empirical Tests

Each period one outcome, 0 or 1, is observed.[3] Let $\Omega = \{0,1\}^{\infty}$ be the set of all *paths*, i.e., infinite histories. A finite history $s_m \in \{0,1\}^m$, $m \geq t$, is a *extension* of $s_t \in \{0,1\}^t$ if the first t outcomes of s_m coincide with s_t. Analogously, a path $s \in \Omega$ extends s_t if $s = (s_t, ...)$. In the opposite direction, let $s_m \mid t$ be the first t outcomes of a finite history $s_m \in \{0,1\}^m$, $m \geq t$. Analogously, $s \mid t \ = s_t$ for any path s that extends $s_t \in \{0,1\}^t$. A *cylinder* with base on s_t is the set $C(s_t) \subset \{0,1\}^{\infty}$ of all infinite extensions of s_t. We endow Ω with the topology that comprises of unions of cylinders with finite base. Let $\Im_t$ be the algebra that consists of all finite unions of cylinders with base on $\{0,1\}^t$. Denote by N the set of natural numbers. Let $\Im$ is the σ-algebra generated by the algebra $\Im^0 \equiv \bigcup\limits_{t \in N} \Im_t$, i.e., $\Im$ is the smallest $\sigma-$algebra which contains $\Im^0$.

Let $\Delta(\Omega)$ the set of all probability measures on $(\Omega, \Im)$. We endow $\Delta(\Omega)$ with the weak*-topology and with the $\sigma-$algebra of Borel sets, (i.e., the smallest $\sigma-$algebra which contains all open sets in weak*-topology).[4] Let $\Delta\Delta(\Omega)$ be the set of probability measures on $\Delta(\Omega)$.

Let f be a function

$$f : \bigcup_{t \geq 0} \{0,1\}^t \longrightarrow [0,1]$$

that maps finite histories into next period's forecasts, i.e., a probability of 1 next period.[5] It is well known that f defines a probability measure $P \in \Delta(\Omega)$ on the space of paths. To simplify the language, a probability measure in $\Delta(\Omega)$ is also called a *theory*. That is, a theory is identified with it's predictions. Before any data is

[3] It is immediate to extend the results to the case where there are finitely many possible outcomes in each period.

[4] The weak*-topology consists of all unions of finite intersections of sets of the form

$$\left\{ Q \in \Delta(\Omega) : \left| E^P h - E^Q h \right| < \varepsilon \right\},$$

where $P \in \Delta(\Omega)$, $\varepsilon > 0$, and h is a real-valued and continuous function on Ω. We refer the reader to Rudin (1973), Chapter 3, for this definition and for basic results regarding the weak*-topology.

[5] By convention, $\{0,1\}^0 = \{\emptyset\}$.

observed, Bob announces his theory $P \in \Delta(\Omega)$. A tester, named Alice, tests Bob's theory empirically.[6]

Definition 1. *A test is a function* $T : \Omega \times \Delta(\Omega) \to \{0, 1\}$.

A test is an arbitrary function that takes as an input a theory and a path and returns, as an output, either 0 or 1. When the test returns a 1, the test does not reject the theory based on the path. When a 0 is returned, the theory is rejected.

A test divides paths into those $(A_P) \equiv \{s \in \Omega | \ T(s, P) = 1\}$ where the theory P is not rejected and those $(A_P)^c$ where the theory is rejected. The set A_P is called the *acceptance set*. The set $(A_P)^c$ is called the *rejection set*. We consider only tests T such that the acceptance sets A_P are $\Im-$measurable. Fix $\varepsilon \in [0, 1]$.

Definition 2. *A test T does not reject the truth with probability $1 - \varepsilon$ if for any $P \in \Delta(\Omega)$*

$$P\{s \in \Omega \mid T(s, P) = 1\} \geqslant 1 - \varepsilon.$$

A test does not reject the truth if the actual data generating process is not rejected with high probability. A theory that fails such a test can be (with high confidence) reliably viewed as false.

Many results in probability and statistics take the form

$$P(A_P) = 1 \text{ or, more generally, } P(A_P) \geqslant 1 - \varepsilon.$$

These results (such as the law of large numbers, the law of iterated logarithms, the central limit theorem) relate the unobservable concept of a probability measure $P \in \Delta(\Omega)$ with a (potentially) observable set of paths $A_P \in \Im$.[7] A successful result

[6]It might be worth emphasizing that even though Alice tests Bob's theory using a string of outcomes, we do not make distributional assumptions such as an independent, identically distributed process. These are strong and difficult to validate assumptions. If, before any data was observed, Alice knew that the actual process belongs to a small set such as independent, identically distributed processes then she could infer the actual process from the data, with no need for an informed expert.

[7]Some of these findings were initially demonstrated within a class of theories, but martingale techniques allow these distributional assumptions to be relaxed.

showing that a well defined event occurs almost surely (or with high probability) can be used to construct a test that does not reject the truth. Over several decades, many empirical tests have been developed. Below we show a few well-known examples of them.

2.1. Classic examples of tests

Given a theory $P \in \Delta(\Omega)$, path $s \in \Omega$, and $s_t = s \mid t$, let

$$f_0^P(s) \equiv P(C(1)) \text{ and } f_t^P(s) \equiv \frac{P(C(s_t, 1))}{P(C(s_t))}$$

be forecasts made along s.[8] Denote by $I_t(s)$ the $t-$th coordinate of s. The test

$$T(s, P) = 1 \text{ if and only if } \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} [f_{t-1}^P(s) - I_t(s)] = 0,$$

requires the average forecast of 1 to match the empirical frequency of 1. Lehrer (2001) considered tests that require the match between average forecasts and empirical frequencies to occur on several subsequences such as even periods, odd periods, etc. The calibration test of Foster and Vohra (1998) requires the empirical frequency of 1 to be near $p \in [0, 1]$ in the periods that 1 was forecasted with probability close to p.

An example of a well-known test (not based on matching empirical frequencies) is the likelihood test: Let $a : \Delta(\Omega) \longrightarrow \Delta(\Omega)$ be any function such that $P^a \equiv a(P)$ and P are singular ($P \perp P^a$).[9] The *likelihood test* T is defined by

$$T(s, P) = 1 \text{ if and only if } \lim_{n \to \infty} \frac{P^a(C(s_n))}{P(C(s_n))} = 0, \; s_n = s \mid n.$$

The likelihood test requires the theory P to outperform an alternative theory P^a in the sense that, on the observed path s, the likelihood of P^a becomes arbitrarily smaller than the likelihood of P.

A proof that the calibration and likelihood tests do not reject the truth with probability one can be found in Dawid (1982, 1985).

[8] If $P(C(s_t)) = 0$ then $f_t^P(s)$ is arbitrarily defined.

[9] P and P^a are singular if there is a set $A \in \Im$ such that $P(A) = 1$ and $P^a(A) = 0$.

2.2. Passing tests without knowledge

Consider the uninteresting test such that the acceptance set of a theory $\bar{P}$ is the entire space of paths Ω. This test does not reject $\bar{P}$ on any path. Hence, Bob can pass this test by simply announcing $\bar{P}$. No knowledge over the data generating process is required. More generally, Bob may be able to pass other tests (on all paths) by selecting theories at random.

Definition 3. *A test T can be ignorantly passed with probability $1-\varepsilon$ if there exists a random generator of theories $\zeta_T \in \Delta\Delta(\Omega)$ such that for every path $s \in \Omega$*

$$\zeta_T \{P \in \Delta(\Omega) \mid T(s, P) = 1\} \geqslant 1 - \varepsilon.$$ [10]

The random generator ζ_T may depend on the test T, but not on any knowledge of the actual data generating process. If a test can be ignorantly passed then for any given path $s \in \Omega$, the acceptance set A_P contains s, with ζ_T-probability $1-\varepsilon$. Hence, Bob can randomly select theories such that, with probability $1-\varepsilon$ (according to Bob's randomization device), will not be rejected, no matter which path is realized.

A test that can be ignorantly passed delivers the same verdict to experts who knows the data generating process and to experts who knows nothing about it. Thus, a test that can be ignorantly passed cannot determine whether an expert has relevant knowledge (i.e., knowledge the tester does not have).

Assume that for every random generator of theories $\zeta \in \Delta\Delta(\Omega)$ there exists at least one path s such that, with probability greater than ε, the realized theory is rejected on s. Then, by definition, the test cannot be passed with probability $1-\varepsilon$. A single path in which rejection may occur suffices as evidence that the test cannot be passed. However, a stronger property may be demanded. The tester may be interested in paths such that an uninformed expert fails the test with near certainty (as opposed to positive probability).

[10] This definition requires a measurability provision on the sets $\{P \in \Delta(\Omega) \mid T(s, P) = 1\}$. However, in this paper, either we consider concrete examples of tests and random generators of theories, for which there is no measurability issue or, for general tests, we limit ourselves to random generators of theories with finite support, which automatically resolves the issue of measurability.

Definition 4. *Fix a test* T*. Given a random generator of theories* $\zeta \in \Delta\Delta(\Omega)$ *and* $\varepsilon \geq 0$*, let* $R^{\varepsilon}_{\zeta} \subseteq \Omega$ *be the set of all paths* $s \in \Omega$ *such that*

$$\zeta \{P \in \Delta(\Omega) \mid T(s, P)) = 0\} \geq 1 - \varepsilon.$$

The set R^{ε}_{ζ} is called the *revelation set*, where the random generator of theories ζ fails to manipulate the test with probability $1 - \varepsilon$.

Definition 5. *A test* T *cannot be* $\varepsilon-$*manipulated,* $\varepsilon \geq 0$*, if for every random generator of theories* $\zeta \in \Delta\Delta(\Omega)$*, the revelation set* R^{ε}_{ζ} *is not empty.*[11]

Theory rejection is feasible if the test that cannot be manipulated. No matter which random generator of theories is used, there exists at least of a path that induces, with probability $1 - \varepsilon$, the rejection of the realized theory.

The calibration tests of Foster and Vohra (1998) and Lehrer (2001) are such that for every theory $P \in \Delta(\Omega)$ the rejection set $(A_P)^c$ is not empty. So, for any given theory, there are paths in which the theory is rejected. So, it may seem that theory rejection is viable, but this feasibility is effectively removed because calibration tests can be ignorantly passed. An actual prospect of theory rejection requires a demonstration that the revelation sets are not empty.

Definition 6. *A test is* ε*-effective,* $\varepsilon \geq 0$*, if it does not reject the truth with probability* $1 - \varepsilon$ *and cannot be* $\varepsilon-$*manipulated.*

2.3. Acceptance and Rejection Tests

In general, a test may reject and accept theories at infinity. Hence, some structure is needed to preserve the interpretation of a test as a mechanism of relating the unobservable concept of a theory with potentially observable data.

Definition 7. *Acceptance tests are such that, for all theories* $P \in \Delta(\Omega)$*, the acceptance set* A_P *is a union of cylinders with base on a finite history. Rejection tests are such that, for all theories* $P \in \Delta(\Omega)$*, the rejection set* $(A_P)^c$ *is a union of cylinders with base on a finite history.*

[11] If a test cannot be $0-$manipulated then we simply say that it cannot be manipulated.

An acceptance test stops, at some finite time, when the observed data is deemed consistent enough with the announced theory. A rejection test stops, at some finite time, when the data is regarded as inconsistent enough with the announced theory. A finite test is both a rejection and an acceptance test.

Proposition 1. *Fix* $\varepsilon \in [0,1]$ *and* $\delta \in (0, 1-\varepsilon]$. *Let* T *be an acceptance test that does not reject the truth with probability* $1-\varepsilon$. *Then, the test* T *can be ignorantly passed with probability* $1-\varepsilon-\delta$.

Assume that Alice's test does not reject the truth and accepts Bob's theory only if the data supports the theory (i.e., the observed history belongs to the acceptance set). Then, an uninformed expert can produce theories at random such that, no matter which path is observed, the realized theory will be supported by the data, with high probability. Thus, no relevant knowledge is necessary to produce theories that will, in the future, prove to be supported by the data.[12]

The widespread practice of supplying theories that account for the facts is vulnerable to the usual reproach centered at the existence of alternative theories that also explain the data. Let us say that a theory P is consistent with the data $s_t \in \{0,1\}^t$ if $C(s_t)$ is contained in the acceptance set of a test T. Given s_t, it is straightforward to find multiple theories consistent with it. Proposition 1 shows that, without knowing the data, Bob can fabricate theories that are consistent with it. This result makes the standard critique salient.

By proposition 1, the use of acceptance tests are fairly limited. No acceptance test is effective. We now explore rejection tests.

[12] Proposition 1 is a surprising result. Assume that Alice writes a long sentence of zeros and ones. Bob simulataneously selects a probability measure over all possible sentences. Alice extracts a set A of finite sentences that has high probability according to Bob's choice. She agrees to accept that Bob has some relevant signal if her sentence belongs to A (in this illustration, Alice combines the roles of Nature and tester). Bob knows which set A Alice extracts from any given theory, but he knows nothing about what Alice has written. He can (almost) assure himself that her sentence belongs to his extracted set A.

Definition 8. *A test T_2 is harder than a test T_1 if*

$$\{s \in \Omega |\ T_2(s,P) = 1\} \subseteq \{s \in \Omega |\ T_1(s,P) = 1\}.$$

If T_2 is harder than T_1 then rejection by T_1 implies rejection by T_2.

Proposition 2. *Fix $\delta \in (0,1]$. Let T_1 be any test that does not reject the truth with probability $1-\varepsilon$. There exists a rejection test T_2 that is harder than the test T_1 and does not reject the truth with probability $1-\varepsilon-\delta$.*

Proposition 2 shows that given any test T_1 there exists a rejection test T_2 that does not reject the truth with almost equally high probability as T_1 does, and the rejection sets of T_2 contain those of T_1. Thus, if T_1 has non-empty revelation sets then T_2 is a rejection test with non-empty revelation sets. We refer to T_2 as the rejection test associated with the test T_1.

2.4. Intuition of Proposition 1 and Proof of Proposition 2

Consider the following game zero-sum game between Nature and Bob. Nature chooses a probability measure $P \in \Delta(\Omega)$. Bob chooses a random generator of theories $\zeta \in \Delta\Delta(\Omega)$. Bob's payoff is $E^{\zeta}E^{P}T$, where E^P and E^{ζ} are the expectation operators associated with P and ζ, respectively.

By assumption, the test T does not reject the truth with probability $1-\varepsilon$. Therefore, for every strategy P of Nature, there is a strategy ζ_P for Bob (that assigns probability one to P) such that Bob's payoff

$$E^{\zeta_P}E^{P}T = P\{s \in \Omega \mid T(s,P) = 1\}$$

is greater than $1-\varepsilon$. Hence, if the conditions of Fan's MinMax are satisfied then there is a strategy ζ_T for Bob that ensures Bob a payoff arbitrarily close to $1-\varepsilon$, no matter which strategy Nature chooses. In particular, Nature could use P_s, selecting a single path s with certainty. Therefore, for all $s \in \{0,1\}^{\infty}$,

$$E^{\zeta_T}E^{P_s}T = \zeta_T\{P \in \Delta(\Omega) \mid T(s,P) = 1\} \geq 1-\varepsilon-\delta.$$

Fan's MinMax theorem requires the payoff of one of the players to be lower semi-continuous and the strategy space of that player to be compact. As is well-known, $\Delta(\Omega)$ is compact in the weak*-topology. The lower semi-continuous of Bob's payoff follows from the openness of the acceptance sets. It is here that the assumption of an acceptance test is used. In the case of a rejection test, Bob's payoff function is upper semi-continuous.

Proposition 2 is a direct corollary of the following well-known result: for any given probability measure $P \in \Delta(\Omega)$ and $\delta > 0$, any set $A \in \Im$ can be enlarged to an open set $U \supset A$ such that $P(U) < P(A) + \delta$ (see Ulam's Theorem, 7.1.4 in Dudley (1989)).

3. Category Tests

Category tests can de defined as follows: Given a set A, let $\bar{A}$ be the closure of A. A set A is called *nowhere dense* when the interior of its closure, $\bar{A}$, is empty. A *first category* set is a countable union of nowhere dense sets. A first category set is (topologically) small. The complement of a first category set is (topologically) large.[13]

Category tests are such that, for any theory $P \in \Delta(\Omega)$, the acceptance set A_P is a first category set. Proposition 3 below shows the existence of category tests that do not reject the truth.

Proposition 3. *Fix $\delta \in (0, 1]$. Given any probability measure $P \in \Delta(\Omega)$, there is a closed set $D_P \in \Im$ with empty interior such that*

$$P(D_P) \geq 1 - \delta.$$

In particular, there exists a first category set $D_P^{/}$ such that $P(D_P^{/}) = 1$. Proposition 3 is classic result (see Oxtoby (1980), Theorem 16.5 and Dekel and Feinberg (2006)). It shows that any theory makes at least one sharp prediction: a path in a topologically small set will be realized.

[13] There are others definitions of small sets that we have not examined (see Anderson and Zame (2001) and Stinchcombe (2001)).

Proposition 4. *Dekel and Feinberg (2006). Given any first category set $A \in \Im$, the set of probability measures that assign positive probability to A,*

$$\{P \in \Delta(\Omega) \mid P(A) > 0\}$$

is a a first category subset of $\Delta(\Omega)$.

By proposition 4, any theory will be rejected by a category test for all data generating processes except for a topologically small set of them. A sufficiently close match between the announced theory and the data generating process seems to be a prerequisite to pass a category test. So, one may conjecture that, apart from extraordinary coincidence, a completely uninformed expert find it impossible to pass a category test. However, this does not necessarily hold. We show an example of a category test that can be ignorantly passed with arbitrarily high probability. Topologically large rejection sets may not deliver feasible rejection of false experts.

3.1. Falsifiability

Category tests are defined by topological concepts. A more intuitive way of understanding these tests is to relate them to the celebrated concepts of Popper (1958). Popper is interested in a demarcation criteria that would differentiate scientific ideas from non-scientific ideas (and, hence, would give meaning to the term scientific). He proposes that scientific theories are those that can be falsified (i.e., rejected by the data).

Falsifiability is a central concept in near all scientific disciplines. Here, falsifiability must be understood as the possibility of conclusive empirical rejection. This is not the same as the canonical principle in statistics that the observation of events that were ex-ante deemed unlikely (by the theory) may suffice for it's rejection.

Definition 9. *A theory $P \in \Delta(\Omega)$ is falsifiable if, for every finite history $s_t \in \{0,1\}^t$, there is an extension s_m of s_t such that*

$$P(C(s_m)) = 0. \tag{3.1}$$

So, a theory is falsifiable if, after any finite history, there is a finite sequence of outcomes that bluntly contradicts it.

Let $\digamma \subset \Delta(\Omega)$ be the set of falsifiable theories. Given $P \in \digamma$, let $\left(\overline{A}_P\right)^c$ be the union of all cylinders $C \in \Im^0$ such $P(C) = 0$, and let $\overline{A}_P$ be the complement of $\left(\overline{A}_P\right)^c$, i.e., the paths that do not bluntly contradict a falsifiable theory P.

The *falsifiability test* rejects non-falsifiable theories out of hand (i.e., on all paths). Any falsifiable theory bluntly contradicted by the data is also rejected. Formally, the falsifiability test T is defined by

$$T(s,P) = 1 \qquad \text{if } s \in \bar{A}_P \text{ and } P \in \digamma;$$

$$T(s,P) = 0 \qquad \text{otherwise.}$$

Proposition 5. *Fix* $\varepsilon \in (0,1]$. *The falsifiability test is a category test. Moreover, the falsifiability test can be ignorantly passed with probability* $1-\varepsilon$.

By proposition 5, it is possible to produce falsifiable theories at random such that, with arbitrarily high probability, the realized theory will not be falsified, regardless of which data is eventually observed.

By definition, falsifiability means the possibility of empirical rejection. In itself, this possibility is rendered immaterial because the falsifiability test can be ignorantly passed. The falsifiability criteria are too weak to produce an effective test.

Propositions 4 and 5 show an interesting contrast. Any theory is rejected by the falsifiability test, unless data generating processes belongs to a first category set (that depends on the announced theory). However, Bob, without any knowledge over the actual process, can ignorantly pass the falsifiability test.

Dekel and Feinberg (2006) show that if the continuum hypothesis holds true then there exists a category test that does not reject the truth and cannot be manipulated. This result is an important achievement, but it does not show the existence of an effective test because it relies on the continuum hypothesis. It leaves open the possibility that, like the continuum hypothesis, the existence of their test is an undecidable question with standard axioms.

3.2. Intuition of Proposition 5

It is straightforward to show that the falsifiability test is a category test. The acceptance set of a non-falsifiable theory is empty and, hence, of first category. The rejection set of a falsifiable theory is open (as unions of cylinders with finite base) and, by definition, dense. Hence, the acceptance set of a falsifiable theory is closed and nowhere dense.

The intuition that the falsifiability test can be ignorantly passed is as follows: Consider an increasing sequence of natural numbers Z_t, $t \in N$, and a sequence of independent random variables such that each variable is uniformly distributed over the finite set $X_t \equiv \{0,1\}^{(Z_{t+1}-Z_t)}$. Given $x = (x_t,\ x_t \in X_t)$, let P^x be a falsifiable theory that is rejected if and only if, for some $t \in N$, x_t occurs between periods Z_t and Z_{t+1}. Now assume that Z_t grows sufficiently fast (and Z_1 is sufficiently large) so that the chances that x_t, $t \in N$, be realized (at least) once is small. Let the falsifiable theory P^x be selected if x is selected. Fix any path $s = (y_1, ..., y_t, ...)$, $y_t \in X_t$. The selected theory P^x will be falsified on s if and only if $x_t = y_t$ for some $t \in N$. By construction, this is an unlikely event.

4. An Effective Test

In this section, we construct a test that does not reject the truth and cannot be manipulated. Let $s^1, s^2, ...$ be a countable dense subset of Ω. Fix $k \in N$. For every path s^i, there exists a period $t \in N$ such that the cylinder $C_P^k(s_t^i)$ with base on the finite history s_t^i satisfies

$$P(C_P^k(s_t^i) - \{s^i\}) \leq \frac{1}{2^{k+i}}. \tag{4.1}$$

Indeed, the sequence of sets $C_P^k(s_t^i) - \{s^i\}$ is descending (as t goes to infinity), and its intersection is empty. So, $P(C_P^k(s_t^i) - \{s^i\})$ goes to zero as t goes to infinity.

Let $t(i,k,P)$ be the smallest natural number such that (4.1) is satisfied. We define $A_k^i(P) \equiv \Omega - C_P^k(s_t^i)$ as the complement of $C_P^k(s_t^i)$, where $t = t(i,k,P)$. Let

$$A_k(P) \equiv \bigcap_{i=1}^{\infty} A_k^i(P) \text{ and } \hat{A}_P \equiv \bigcup_{k=1}^{\infty} A_k(P) \cup \bigcup_{i=1}^{\infty} \{s^i\}.$$

Each set $A_k(P)$ is an intersection of closed sets and, hence, it is closed. By construction, each set $A_k(P)$ has an empty interior, and hence, $\hat{A}_P$ is a first category set.

The *global category test* $\hat{T}$ is such that $\hat{A}_P$ is the acceptance set of $P \in \Delta(\Omega)$. That is,

$$\hat{T}(s, P) = 1 \qquad \text{if } s \in \hat{A}_P;$$

$$\hat{T}(s, P) = 0 \qquad \text{if } s \notin \hat{A}_P.$$

Let $\hat{R}^0_\zeta$ be the revelation set where the random generator of theories $\zeta \in \Delta\Delta(\Omega)$ fails to manipulate the global category test with probability one.

Proposition 6. *The global category test $\hat{T}$ does not reject the truth with probability one and it cannot be manipulated. Moreover, given any random generator of theories $\zeta \in \Delta\Delta(\Omega)$, the revelation set $\hat{R}^0_\zeta$ is such that $P(\hat{R}^0_\zeta) = 1$ for every probability measure $P \in \Delta(\Omega)$, except for, at most, a first category set of them.*

The global category test controls the type I error of rejecting an informed expert and the type II error of accepting an uniformed expert. The informed expert passes the global category test with probability one. The uninformed expert can employ any random generator of theories ζ, but failure is inevitable on the paths of the revelation set $\hat{R}^0_\zeta$. This set is large enough so that all probability measures, except for a first category set, assign probability one to it.

The tests proposed in the literature are strikingly different from the global category test. In contrast to the preceding tests, the global category test can discredit an uniformed expert. Moreover, it is not only feasible to reject an uninformed expert. It is also plausible given that, no matter how Bob randomizes, rejection is near certain outside a first category set of data generating processes.

Given $\varepsilon > 0$, let $\hat{T}^\varepsilon$ be a rejection test, associated with the global category test $\hat{T}$, that does not reject the truth with probability $1-\varepsilon$. Corollary 1, below, is immediate from propositions 2 and 6.

Corollary 1. *The rejection test* $\hat{T}^{\varepsilon}$ *does not reject the truth with probability* $1-\varepsilon$ *and it cannot be manipulated. The revelation sets of* $\hat{T}^{\varepsilon}$ *contain the revelation sets of the global category test and, hence, also have full measure for all probability measures* $P \in \Delta(\Omega)$, *except for, at most, a first category set of them.*

Corollary 1 shows the existence of an effective rejection test $\hat{T}^{\varepsilon}$. Hence, for any given theory P, the test $\hat{T}^{\varepsilon}$ demarcates the rejection set of P as the set of finite histories of outcomes that must be regarded as sufficiently inconsistent with the theory P to validate it's rejection. So, if an outcome sequence in the rejection set of the theory P is observed then P must be rejected. This result delivers an empirical test that takes into account that theories might be strategically produced. If Alice adopts the test $\hat{T}^{\varepsilon}$ then false experts can be discredited.

In addition, corollary 1, in conjunction with proposition 1, shows a basic asymmetry between acceptance and rejection tests. Every acceptance test that does not reject the truth can be ignorantly passed, but there exists a rejection test that does not reject the truth and cannot be manipulated. So, propositions 1, 2, and 6, combined, deliver a formal support to the maxim that testable theories must be accepted until proven inconsistent with the data rather than rejected until proven consistent with the data.

Proposition 6 removes the conceptual bound that would be imposed on the significance of empirical testing of theories if an effective test did not exist. An effective test has a clear purpose. It does not reject a theory that coincides with the data generating process and it may, for some realizations of the data, reject the false theories of an uninformed expert. Thus, a rejection verdict from an effective test is feasible and the rejected theories are, almost surely, false.

4.1. Intuition of Proposition 6

The proof that the global category test does not reject the truth follows from the fact that, for any $k \in N$, the rejection set of a theory P is contained in the union $i = 1, ...$ of all cylinders $C^k_P(s^i_t)$ (minus the paths s^i). By (4.1), the probability of this union is less than 2^{-k}. Given that this holds for any $k \in N$, the probability of any rejection set is zero.

The intuition for the proof that the global category test cannot be manipulated is follows: Each acceptance set $\hat{A}_P$ is a first category set that has $P-$full measure. So, once Bob proposes a theory, he passes the test if the observed path belongs to a topologically small acceptance set and is rejected on a topologically large rejection set. This creates a difficulty for an uninformed expert. By itself, this difficulty is not unsurpassable as demonstrated by proposition 5. The main idea is that, in a global category test, all acceptance sets (for different theories P) are constructed around the same paths $\{s^i\}$. Acceptance sets of different theories are sufficiently close to each other so that, for any random generator of theories ζ, the *union* of $\zeta-$almost all acceptance sets is a first category set of paths. By definition, failure is unavoidable outside this union. So, by proposition 4, except for a topologically small set, this union has probability zero for all probability measures.

5. Delaying Rejection

A natural concern with proposition 6 is that the revelation sets $\hat{R}^0_\zeta$ might not contain any cylinders with finite base. However, it follows from the proof of proposition 6 that for every $\varepsilon > 0$ and $\zeta \in \Delta\Delta(\Omega)$, the revelation set $\hat{R}^\varepsilon_\zeta$, where ζ fails to manipulate the global category test with probability $1 - \varepsilon$, contains an open and dense set (by proposition 4, this set has probability one for all probability measures, except for, at most, a first category set of them). If a path on this set is realized then, with probability $1 - \varepsilon$, in some finite time, the global category test fails an uninformed expert. This result, however, does not determine the exact periods that an uninformed expert fails.

Definition 10. *A test T rejects a theory $P \in \Delta(\Omega)$ on a finite history $s_t \in \{0,1\}^t$ (denoted $T(s_t, P) = 0$) if $T(s, P) = 0$ for all paths s that extend s_t.*

So, a test rejects a theory on a finite history s_t if it rejects the theory on all paths s such that the first t outcomes of s are s_t.

Definition 11. *Rejection by a test T can be delayed by m periods with probability $1-\varepsilon$ if there exists a random generator of theories $\zeta_{T,m} \in \Delta\Delta(\Omega)$ such that for every path $s_m \in \{0,1\}^m$,*

$$\zeta_{T,m}\{P \in \Delta(\Omega) \mid T(s_m, P) = 1\} \geq 1 - \varepsilon.$$

Rejection by a test T can be arbitrarily delayed with probability $1-\varepsilon$ if it can be delayed by m periods, with probability $1-\varepsilon$, for every $m \in N$.

If rejection can be arbitrarily delayed then Bob can first choose an arbitrary period m and then randomly select theories such that, with high probability (according to Bob's randomization), will not be rejected before period m, no matter which path is realized.

Proposition 7. *Fix $\varepsilon \in [0,1]$ and $\delta \in (0, 1-\varepsilon]$. Let T be an arbitrary test that does not reject the truth with probability $1-\varepsilon$. Then, rejection by the test T can arbitrarily delayed with probability $1-\varepsilon-\delta$.*

Proposition 7 shows that Bob can maintain a false reputation of being a knowledgeable expert for an arbitrarily long time horizon. This result holds because any test can be approximated by finite tests (a corollary of proposition 2) and, by proposition 1, any finite test that does not reject the truth can be ignorantly passed.

There is an interesting contrast between propositions 6 and 7. Assume that Alice tests Bob with the rejection test associated with the global category test and Bob is not knowledgeable. By proposition 7, Bob can use random generator of theories $\hat{\zeta}_m$ to delay rejection for m periods. A limit, $\hat{\zeta}$, of (a subsequence) of these random generators of theories exists (because $\Delta\Delta(\Omega)$ is compact in the weak*-topology). By propositions 2 and 6, $\hat{\zeta}$ does not delay rejection indefinitely, with strictly positive probability. Moreover, since Alice uses a rejection test, given any $\delta > 0$, $\hat{\zeta}$ does not delay rejection for m periods, with probability δ, if m is sufficiently large. So, a change from $\hat{\zeta}_m$ to $\hat{\zeta}$ (that becomes arbitrarily small as m increases) triggers an abrupt transformation: rejection delay, that was formerly assured with arbitrarily high probability, is no longer assured even with small probability. Therefore, to delay rejection for m periods (m large) Bob must choose the random generator of theories in a very precise way.

6. Rejecting Theories Out of Hand

Proposition 7 shows that the class of permissible theories must be restricted. This does follow from a need to simplify the analysis, but rather because in the absence of such restrictions a completely ignorant expert can delay rejection for arbitrarily many periods, on all future realizations of the data. This difficulty is hard to overcome unless we relax the main assumption of proposition 7: that test that does not reject the truth.[14] So, we allow the test to reject some theories out of hand (i.e., on all paths). This is clearly not a desiderata because potentially correct theories are rejected by definition. Even so, the test may remain powerless to reject a false expert with bounded data sets.

Definition 12. *A test T does not reject an informed expert with probability $1-\varepsilon$ if for every theory $P \in \Delta(\Omega)$ there exists a theory $\tilde{P} \in \Delta(\Omega)$ such that*

$$P\left\{s \in \Omega \mid T(s, \tilde{P}) = 1\right\} \geqslant 1 - \varepsilon.$$

The test T does not reject an informed expert if Bob, knowing that the data generating process is P, can announce a theory $\tilde{P}$ that passes the test. Clearly, if a test that does not reject the truth then it does not reject an informed expert, but the converse is not necessarily true.

The falsifiability test rejects the truth, but does not reject an informed expert with arbitrarily high probability. Fix $\delta > 0$. Assume that an informed expert knows that the data generating process is a non-falsifiable theory P. Let D_P be a closed set with empty interior such that $P(D_P) \geq 1 - \delta$ (see proposition 3 for the existence of D_P). Let $\tilde{P}$ denote P conditional on D_P. That is,

$$\tilde{P}(B) = \frac{P(B \cap D_P)}{P(D_P)} \text{ for all } B \in \Im.$$

[14] An alternative, not taken in this paper, is to impose complexity bounds on Bob's strategies.

By lemma 2 (in the appendix), $\tilde{P}$ is a falsifiable theory and, by definition, any set $D_{\tilde{P}}$ that has full $\tilde{P}-$measure, has $P-$probability greater than $1-\delta$. So, with arbitrarily high probability, an informed expert can pass the falsifiability test by announcing $\tilde{P}$.

Consider any dense set $\Lambda \subseteq \Delta(\Omega)$ of probability measures in the topology induced by the sup-norm.[15] Let T be any test that does not reject the truth with probability $1-\varepsilon$. Let T' be the test that coincides with T on Λ, but rejects out of hand theories outside Λ. That is, $T'(s,P) = T(s,P)$ if $P \in \Lambda$ and $T'(s,P) = 0$ if $P \in (\Lambda)^c$. Then, T' does not reject an informed expert with probability $1-\varepsilon-\delta$ for every $\delta \in (0, 1-\varepsilon]$.

Propositions 1 and 7 still hold if the assumption that the test does not reject the truth with probability $1-\varepsilon$ is replaced by the weaker assumption that the test does not reject an informed expert with probability $1-\varepsilon$ (the proofs in the appendix make this weaker assumption). Hence, an uniformed expert can still, with high probability, arbitrarily delay rejection by the test T' (that rejects out of hand all theories outside Λ). In particular, rejection can be arbitrarily delayed if T' rejects out of hand finitely many theories.

We now consider a smaller set of permissible theories. Meteorological forecasts are usually evaluated by how low their Brier Score is (see Brier (1950)).[16] A low Brier Score is not possible if the forecasts are often close to 0.5. Given $\delta > 0$ and $m \in N$, let $\Lambda_{m,\delta} \subseteq \Delta(\Omega)$ be the set of theories P such that for all $s \in \Omega$

$$\frac{1}{m}\sum_{t=1}^{m}(f_{t-1}^{P}(s) - 0.5)^2 > \delta.$$

[15] The sup-norm is such that given two theories P and $P^{/}$,

$$\left\| P - P^{/} \right\| = \sup_{A \in \Im} \left| P(A) - P^{/}(A) \right|.$$

[16] The Brier Score of a theory P at period m is

$$\frac{1}{m}\sum_{t=1}^{m}[f_{t-1}^{P}(s) - I_t(s)]^2.$$

The set $\Lambda_{m,\delta}$ excludes theories that forecast 1 and 0 with near equal odds, sufficiently often, until period m.

Fix $\delta > 0$ and $b \in (0,\, 0.5)$. Given any theory P, let $\bar{P}$ be the theory such that: for every $s \in \Omega$ and $t \in N$, $f_t^{\bar{P}}(s) =$

$$\begin{array}{cc} \frac{1}{2}+b & \text{if } f_t^P(s) \in (\frac{1}{2}+\frac{\delta}{2}, 1]; \\ \frac{1}{2}-b & \text{if } f_t^P(s) \in [0, \frac{1}{2}-\frac{\delta}{2}); \\ \frac{1}{2} & \text{otherwise.} \end{array}$$

That is, given a theory P, the theory $\bar{P}$ forecast 1 with probability: $\frac{1}{2}+b$ when the P forecast (of 1) is high; $\frac{1}{2}$ when the P forecast is intermediate; and $\frac{1}{2}-b$ when the P forecast is low.

Let $\beta \equiv (\delta,\, m, b,\, K)$, where $K \in N$. Let T^{β} be the test

$$T^{\beta}(s,P) = \begin{array}{cc} 1 & \text{if } \bar{P}(C(s_m)) \geq K(0.5)^m,\ s_m = s \mid m; \\ 0 & \text{otherwise.} \end{array}$$

The test T^{β} is akin, but not identical, to a calibration test. Define a success when the observed outcome was forecasted with probability greater than $0.5(1+\delta)$. Define a failure when the observed outcome was forecasted with probability smaller than $0.5(1-\delta)$. The test T^{β} requires the frequency of success to be greater than the frequency of failure (by a factor that depends on b, m, and K).[17]

The test T^{β} rejects (or not) any theory at period m. So, by definition, T^{β} cannot be $\varepsilon-$manipulated if and only if rejection by T^{β} cannot be delayed, with probability ε, by m periods.

Proposition 8. *For any $\varepsilon > 0$, $\delta > 0$ and $K \in N$, there exist $b > 0$ and $\bar{m} \in N$ such that if $m \geq \bar{m}$ then T^{β} does not reject any theory $P \in \Lambda_{m,\delta}$ with probability $1-\varepsilon$. Moreover, if $K\varepsilon > 1$ then rejection by the test T^{β} cannot be delayed, with probability ε, by m periods.*

[17] See Seidenfeld (1985) for a decomposition of the Brier Score into a calibration test related to T^{β} and a refinement term that is maximized when the empirical frequency of 1 is 0.5.

By proposition 8, the test T^β cannot be manipulated even with limited data. So, feasible rejection by T^β cannot be arbitrarily delayed. Moreover, it is improbable that T^β rejects a theory that coincides with the data generating process, provided that it belongs to the domain of permissible theories.

We now show a formal sense in which the restricted set of allowed theories $\Lambda_{m,\delta}$ can be considered a large set.[18]

Proposition 9. *The set $\Lambda_{m,\delta}$ is an open subset of $\Delta(\Omega)$ for any $\delta > 0$ and $m \in N$. If $\delta \in (0, 0.25)$ then given any theory $P \in \Delta(\Omega)$ and a neighborhood U of P, there exist $\hat{m} \in N$ and a theory Q such that $Q \in U \cap \Lambda_{m,\delta}$ for every $m \geq \hat{m}$.*[19]

So, by proposition 9, $\Lambda_{m,\delta}$ is an open set and, any open set intersects $\Lambda_{m,\delta}$ if m is large enough.

We hope that these results provide the groundwork for future advances on the capability of empirical tests with bounded data sets. However, our results leave open several relevant issues. Among them is the matter of the minimum size of the data required to make the test non-manipulable, how plausible is the dismissal of an uninformed expert, which tests minimize the need for restrictions on the set of permissible theories and which class of theories should be rejected out of hand. Answers to these questions are beyond the scope of this paper.

6.1. Intuition of Proposition 8

The proof that T^β does not reject any theory $P \in \Lambda_{m,\delta}$ with high probability is a straightforward application of the law of large numbers. An outcome tend to occur more often when it is correctly forecasted with high probability than when it is correctly forecasted with low probability. The proof that T^β cannot be $\varepsilon-$manipulated

[18] The set $\Lambda_{m,\delta}$ is not topologically large in the sense that its complement is a first category set. We can show that any set with similar properties to those of $\Lambda_{m,\delta}$ is not large in this sense. Proposition 9 asserts that $\Lambda_{m,\delta}$ is topologically large in a slightly weaker sense.

[19] The theory Q makes the same forecasts as P for several periods and subsequently forecast 1 with high probability.

(when $K\varepsilon > 1$) is also simple. At period m, there are 2^m possible histories. A theory P is rejected unless $\bar{P}$ assigns sufficiently high probability to the observed history. So, rejection occurs in all paths, except for, at most, $2^m/K$ of them. Hence, the following analogy matches Bob's problem well: Bob faces an urn with many balls. Nature picks a ball with unknown odds. Before the chosen ball is revealed, Bob selects, perhaps at random, a small fraction of all possible balls. Then, no matter how Bob randomizes, Bob cannot be near certain that the ball chosen by Nature belongs to his selected set. It is remarkable that, to be properly formalized into a non-manipulable test, such easy and compelling intuition requires a sizeable set of theories rejected out of hand.

7. Conclusion

If an empirical examiner plans to set aside false theories then it must be feasible to reject theories produced by false experts who have no relevant knowledge. Empirical tests with this property are non-manipulable tests. If a rejection is to be a reliable signal that the theory makes incorrect predictions then a theory that forecasts exactly like the data generating process must be rejected with low probability. Empirical tests with this property do not reject the truth.

There is a conflict between these two properties. Empirical tests put forward in the literature do not reject the truth, but can be manipulated. Extensive classes of empirical tests, like acceptance tests, can also be manipulated if they do not reject the truth. One test, the global category test, does not reject the truth and cannot be manipulated. However, any test that does not reject the truth is susceptible to strategic manipulation for arbitrary long periods of time.

An option is to reject some theories out of hand. With this proviso, some tests, akin to those used in practice, do not reject allowed theories (with high probability) and cannot be manipulated. We hope that this result will motivate future advances in the subject matter of strategic manipulation of empirical tests under (perhaps unavoidably) restrictions on the class of permissible theories.

8. Proofs

To prove proposition 1 we need to introduce some terminology and auxiliary lemmas. Let X be a metric space. Recall that a function $f : X \to R$ is *lower semi-continuous* at an $x \in X$ if for every sequence $(x_n)_{n=1}^{\infty}$ converging to x:

$$\forall_{\varepsilon>0} \quad \exists_N \quad \forall_{n\geq N} \quad f(x_n) > f(x) - \varepsilon.$$

The function f is lower semi-continuous if it is lower semi-continuous at every $x \in X$.[20]

Lemma 1. *Let $U \subset X$ be an open set where X is a compact metric space. Equip X with the $\sigma-$algebra of Borel subsets. Let $\Delta(X)$ be the set of all probability measures on X. Equip $\Delta(X)$ with the weak*-topology. The function $F : \Delta(X) \to [0,1]$ defined by*

$$F(P) = P(U)$$

is lower semi-continuous.

Proof: Take a probability measure $P \in \Delta(X)$; we will show that the function F is lower semi-continuous at P. To this end take a sequence $P_n \to_n P$ and an $\varepsilon > 0$; consider a closed set $A \subset U$ such that

$$P(A) > P(U) - \varepsilon/2,$$

and a continuous function $g : \Omega \to [0,1]$ such that

$$\forall_{s\in A} \quad g(s) = 1 \text{ and } \forall_{s\notin U} \quad g(s) = 0;$$

such a set A exists as every open set in a metric space can be represented as the union of an ascending sequence of closed sets, and such a function g exists by the Urysohn Lemma (see, for example, Engelking (1989), Theorem 1.5.11).

[20] See Engelking (1989), Problem 1.7.14 for these definitions and some basic results regarding lower semi-continuous functions.

By the definition of weak*-topology, there exists an N such that for every $n \geq N$

$$\left|E^{P_n}g - E^P g\right| < \varepsilon/2.$$

Thus, if $n \geq N$, then

$$F(P_n) = P_n(U) \geq E^{P_n}g \geq E^P g - \varepsilon/2 \geq P(A) - \varepsilon/2 > P(U) - \varepsilon/2 - \varepsilon/2 = F(P) - \varepsilon.$$

■

Theorem (Fan (1953) Let X be a compact and Hausdorff, linear space and Y a linear space (not necessarily topologized).[21] Let f be a real-valued function on $X \times Y$ such that for every $y \in Y$, $f(x, y)$ is lower semi-continuous on X. If f is convex on X and concave on Y, then

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

Proof of the Proposition 1: We prove proposition 1 under a weaker assumption that the test T does not reject an informed expert with probability $1 - \varepsilon$. Let X be $\Delta(\Omega)$. Let Y be the subset of $\Delta(\Delta(\Omega))$ of all random generator of theories with finite support. So, an element ζ of Y can be described by a finite sequence of probability measures $\{P_1, ..., P_n\}$ and positive weights $\{\pi_1, ..., \pi_n\}$ that add to one (i.e., ζ selects P_i with probability π_i, $i = 1, ..., n$). Let the function $f : X \times Y \to R$ defined by

$$f(P, \zeta) \equiv E^{\zeta} E^P T = \sum_{i=1}^{n} \pi_i \int T(s, P_i) dP(s). \tag{8.1}$$

We now check that the assumptions of Fan's theorem are satisfied. Since T is an acceptance test, the set

$$U_Q = \{s \in \Omega : T(s, Q) = 1\}$$

is open for every $Q \in \Delta(\Omega)$. Therefore, by lemma 1,

$$P(U_Q) = \int T(s, Q) dP(s)$$

[21] Fan does not assume that X and Y are linear spaces. We, however, apply his theorem only to linear spaces.

is a lower semi-continuous function of P. Thus, for every $\zeta \in Y$, the function $f(P,\zeta)$ is lower semi-continuous on X as a weighted average of lower semi-continuous functions.

By definition, f is linear on X and Y, and so it is convex on X and concave on Y. By the Riesz and Banach-Alaoglu Theorems, X is a compact space in weak*-topology; it is a metric space, and so Hausdorff, (see for example Rudin (1973), Theorem 3.17).

Thus, by Fan's Theorem,

$$\min_{P\in\Delta(\Omega)} \sup_{\zeta\in\Delta(\Delta(\Omega))} E^{\zeta}E^{P}T = \sup_{\zeta\in\Delta(\Delta(\Omega))} \min_{P\in\Delta(\Omega)} E^{\zeta}E^{P}T.$$

Notice that the left-hand side of this equality exceeds $1-\varepsilon$, as the test T is assumed not to reject an informed expert with probability $1-\varepsilon$; indeed, for a given $P \in \Delta(\Omega)$, take ζ such that $\zeta(\{\tilde{P}\}) = 1$. Therefore the right-hand side exceeds $1-\varepsilon$, which yield the existence of a random generator of theories $\zeta \in \Delta(\Delta(\Omega))$ such that

$$E^{\zeta}E^{P}T > 1-\varepsilon-\delta$$

for every $P \in \Delta(\Omega)$. Taking, for any given $s \in \Omega$, the probability measure P such that $P(\{s\}) = 1$, we obtain

$$\zeta(\{Q \in \Delta(\Omega) : T(s,Q) = 1\}) > 1-\varepsilon-\delta.$$

■

Lemma 2. *If a theory $P \in \Delta(\Omega)$ is falsifiable then $\bar{A}_P \in \Im$ is a closed, full measure set with empty interior. If there exists a closed set $\bar{A} \in \Im$ with empty interior such that $P(\bar{A}) = 1$ then P is falsifiable.*

Proof : Suppose first that P is falsifiable. The set $(\overline{A}_P)^c$ is open as a union of open sets (cylinders). By definition of falsifiability and the topology in Ω, the set $(\overline{A}_P)^c$ is dense. Finally, since the set of all cylinders is countable, and $(\overline{A}_P)^c$ is a union of cylinders C such that $P(C) = 0$, $(\overline{A}_P)^c$ is a set of measure 0. Therefore, its complement $\overline{A}_P$ is a closed, full measure set with empty interior.

Suppose now that there exists a closed set $\overline{A}$ with empty interior such that $P(\overline{A}) = 1$. Since the complement $\overline{A}^c$ of $\overline{A}$ is a dense set, $\overline{A}^c \cap C(s_t) \neq \emptyset$ for every finite history

$s_t \in \{0,1\}^t$, and since $\overline{A}^c$ is an open set, so is $\overline{A}^c \cap C(s_t)$. Therefore there exists a cylinder $C \subset \overline{A}^c \cap C(s_t)$. Since $C \subset C(s_t)$, $C = C(s_m)$ for some extension s_m of history s_t, and since $C \subset \overline{A}^c$, $P(C) \leq P(\overline{A}^c) = 0$.■

Proof of Proposition 5: Let T be the falsifiability test. It is immediate from lemma 2 that T is a category test. We now show that T can be ignorantly passed with probability $1-\varepsilon, \varepsilon > 0$. Without loss of generality assume that $\varepsilon \in (0, 0.5)$. Let $r > 0$ be small enough so that

$$\sum_{t \in N} r^t < \varepsilon.$$

Let $\{M_t,\ t \in N\}$ be a sequence natural numbers such that

$$\frac{1}{2^{M_t}} < r^t.$$

Let $\hat{P} \in \Delta(\Omega)$ be the probability measure such that both 1 and 0 have equal odds in all periods. It will be convenient to denote by X_t the set $\{0,1\}^{M_t}$ and by X the Cartesian product $\prod_{t \in N} X_t$ of sets X_t. It is also convenient to consider a sequence of independent random variables $\tilde{X}_t$, $t \in N$, uniformly distributed on the set X_t. Let $\tilde{X}$ be the random variable $\prod_{t \in N} \tilde{X}_t$, distributed on the set $X = \Omega$, such that $\tilde{X} = x$ if and only if $\tilde{X}_t = x_t$ for all $t \in N$; thus, $\hat{P}$ is the probability distribution of $\tilde{X}$.

Let

$$Z_t \equiv \sum_{j=1}^{t} M_j.$$

Given an $x \in X$, $x = (x_1, ..., x_t, ...)$, $x_t \in X_t$, let

$$C_x \equiv C(x_1) \cup \bigcup_{t \in N} \bigcup_{z_t \in \{0,1\}^{Z_t}} C(z_t, x_{t+1})$$

be the union of the cylinder with base on $x_1 \in X_1$ and the cylinders with base on finite histories of the form (z_t, x_{t+1}), $t \in N$, where z_t is an arbitrary element of $\{0,1\}^{Z_t}$.

Given that x_1 a sequence of M_1 outcomes and x_{t+1} is a sequence of M_{t+1} outcomes, it follows that

$$\hat{P}\Big(\bigcup_{z_t \in \{0,1\}^{Z_t}} C(z_t, x_{t+1})\Big) = \frac{1}{2^{M_{t+1}}} \text{ and } \hat{P}(C(x_1)) = \frac{1}{2^{M_1}};$$

hence,

$$\hat{P}(C_x) \leq \sum_{t \in N} \frac{1}{2^{M_t}} < \varepsilon < 1. \tag{8.2}$$

Let $(C_x)^c$ be the complement of C_x, $x \in X$, and let $\hat{P}^x$ be the conditional probability of $\hat{P}$ on $(C_x)^c$, $x \in X$. That is,

$$\hat{P}^x(A) = \frac{\hat{P}(A \cap (C_x)^c)}{\hat{P}((C_x)^c)} \text{ for all } A \in \Im.$$

Step 1: We now show that $\hat{P}^x$ is falsifiable.

Indeed, the set $(C_x)^c$ is closed (the complement of an open set) and has $\hat{P}^x$ full measure. For any cylinder $C(s_t)$, $s_t \in \{0,1\}^t$, take a $Z_m \geq t$ and an extension $z_m \in \{0,1\}^{Z_m}$ of s_t. Then, by definition, $C(z_m, x_{m+1}) \subset C_x$ and $C(z_m, x_{m+1}) \subset C(s_t)$. Hence, the set $C(s_t) \cap C_x$ is non-empty. This means that $(C_x)^c$ contains no cylinder, or, in other words, it has empty interior. By lemma 2, $\hat{P}^x$ is falsifiable.

Step 2: Let $C(s_n)$, $s_n \in \{0,1\}^n$, be a cylinder not contained in C_x. We show that

$$\hat{P}(C(s_n) \cap (C_x)^c) > 0.$$

Let $\hat{P}_{s_n}$ denote $\hat{P}$ conditional on s_n. By Bayes' rule,

$$\hat{P}(C(s_n) \cap (C_x)^c) = \hat{P}_{s_n}((C_x)^c)\hat{P}((C_x)^c).$$

By (8.2),

$$\hat{P}((C_x)^c) > 0.$$

Let $\hat{C}_x$ be the finite union of all cylinders in C_x with base on a history of length smaller or equal to n. So, $\hat{C}_x$ is a finite union of cylinders $C(r_m) \subset C_x$, where $m \leq n$. The history s_n cannot be an extension of any of the histories $\{(z_t, x_{t+1}) \mid$

$z_t \in \{0,1\}^{Z_t}, t \in N\}$ or x_1. Otherwise, $C(s_n) \subset C_x$. Thus, $C(r_m) \cap C(s_n) = \varnothing$ if $C(r_m) \subset C_x$ and $m \leq n$. So,

$$\hat{P}_{s_n}\left(\hat{C}_x\right) = 0.$$

Let $\bar{C}_x$ be the union of all cylinders in C_x with base on a history of length strictly greater than n. So, if $k = 0, 1, ...$ is the smallest number such that $n < Z_{k+1}$, then $\bar{C}_x$ is the union of the sets $C(z_t, x_{t+1})$, $t \geq k$ and $z_t \in \{0,1\}^{Z_t}$ (and the set $C(x_1)$ if $k = 0$). Suppose first that that $k = 0$. Then, $\hat{P}_{s_n}$ assigns probability either $0.5^{Z_1 - n}$ or 0 to the cylinder $C(x_1)$ (depending on whether x_1 is an extension of s_n). Moreover, $\hat{P}_{s_n}$ assigns probability $\frac{1}{2^{M_{t+1}}}$ to the union of $C(z_t, x_{t+1})$, $z_t \in \{0,1\}^{Z_t}$. Hence,

$$\hat{P}_{s_n}(\bar{C}_x) \leq 0.5 + \sum_{t>1} \frac{1}{2^{M_t}} < 0.5 + \varepsilon < 1. \tag{8.3}$$

By analogous argument, (8.3) holds with $\sum_{t>k}$ replacing $\sum_{t>1}$ for $k > 0$. Obviously, $C_x = \hat{C}_x \cup \bar{C}_x$, and so

$$\hat{P}_{s_n}(C_x) \leq \hat{P}_{s_n}(\hat{C}_x) + \hat{P}_{s_n}(\bar{C}_x) < 1.$$

Thus,

$$\hat{P}_{s_n}((C_x)^c) > 0.$$

Step 3: We show that C_x is the union of all cylinders $C \in \Im^0$ such that $\hat{P}^x(C) = 0$.

Let $C \in \Im^0$ be an arbitrary cylinder. If $\hat{P}^x(C) = 0$ then $\hat{P}(C \cap (C_x)^c) = 0$. By Step 2, C is contained in C_x. On the other hand, if C is contained in C_x then $C \cap (C_x)^c = \varnothing$. Hence, $\hat{P}^x(C) = 0$.

It follows from Steps 1 and 3 that

$$T(s, \hat{P}^x) = \begin{matrix} 0 & \text{if } s \in C_x; \\ 1 & \text{if } s \notin C_x. \end{matrix}$$

Fix a path $s \in \Omega$. Let s be $(\bar{s}_1, ..., \bar{s}_{t.}, ...)$ where $\bar{s}_t \in X_t$. So, $\bar{s}_1$ are the first M_1 outcomes of s and $\bar{s}_{t+1}$ are the M_{t+1} outcomes that follow the first Z_t outcomes of s.

By definition, $s \in C_x$ if and only if $\bar{s}_t = x_t$ for some $t \in N$. Hence,

$$\left\{ x \in X \mid T(s, \hat{P}^x) = 0 \right\} = \{ x \in X \mid x_t = \bar{s}_t \text{ for some } t \in N \}. \tag{8.4}$$

Moreover,

$$\hat{P} \left\{ \tilde{X}_t = s_t \text{ for some } t \in N \right\} \leq \sum_{t \in N} \hat{P} \left\{ \tilde{X}_t = \bar{s}_t \right\} = \sum_{t \in N} \frac{1}{2^{M_t}} \leq \sum_{t \in N} r^t < \varepsilon. \tag{8.5}$$

Let $\zeta \in \Delta(\Delta(\Omega))$ be such that $\hat{P}^x$ is selected whenever $\tilde{X} = x$. It follows from (8.4) and (8.5) that

$$\zeta \left\{ \hat{P}^x \in \Delta(\Omega) \mid T(s, \hat{P}^x) = 0 \right\} < \varepsilon.$$

■

Proof of Proposition 6: Since

$$\Omega - \hat{A}_P \subset \Omega - \left(A_k(P) \cup \bigcup_{i=1}^{\infty} \{s^i\} \right) \subset \bigcup_{i=1}^{\infty} \left[C_P^k(s_t^i) - \{s^i\} \right],$$

for all $k \in N$,

$$P(\Omega - \hat{A}_P) \leq \sum_{i=1}^{\infty} P(C_P^k(s_t^i) - \{s^i\}) \leq \sum_{i=1}^{\infty} \frac{1}{2^{k+i}} = \frac{1}{2^k},$$

also for all $k \in N$, which yields that $P(\hat{A}_P) = 1$.[22] Thus, the global category test does not reject the truth with probability one. It remains to show that the test cannot be

[22]Notice that our construction proves Proposition 3. Since $P(\hat{A}_P) = 1$ and

$$\hat{A}_P = \bigcup_{k=1}^{\infty} A_k(P) \cup \bigcup_{i=1}^{\infty} \{s^i\},$$

for every $\delta \in (0,1)$ there must exist an m such that

$$P \left(\bigcup_{k=1}^{m} A_k(P) \cup \bigcup_{i=1}^{m} \{s^i\} \right) \geq 1 - \delta,$$

and by construction,

$$\bigcup_{k=1}^{m} A_k(P) \cup \bigcup_{i=1}^{m} \{s^i\}$$

is a closed set with empty interior.

manipulated. Suppose we are given a $\zeta \in \Delta\Delta(\Omega)$. We first show that there exist a subset $\hat{A}$ of Ω, which is a countable union of closed sets with empty interior, and a Borel set $B \subset \Delta(\Omega)$ such that:

$$\zeta\,(B) = 1, \tag{8.6}$$

$$\forall_{P \in B} \quad \hat{A}_P \subset \hat{A}. \tag{8.7}$$

We show later (in corollary 2, following this proof) that for every s the set $\{P \in \Delta(\Omega) : s \in \hat{A}_P\}$ is Borel. Since

$$B \subset \{P \in \Delta(\Omega) : s \in \hat{A}_P\}$$

for every $s \in \Omega - \hat{A}$, we will obtain that

$$\forall_{s \in \Omega - \hat{A}} \quad \zeta\left(\{P \in \Delta(\Omega) : s \in \hat{A}_P\}\right) = 0.$$

This means that $\Omega - \hat{A} \subset R^0_\zeta$, and so the complement of R^0_ζ is a first category set. To complete the proof one has to apply proposition 4.

We will now construct sets $\hat{A}$ and B with properties (8.6)-(8.7). Consider the sets

$$B^i_k(m) \equiv \{P \in \Delta(\Omega) : t(i,k,P) > m\}\,;$$

lemma 3, following this proof, shows that the sets $B^i_k(m)$ are open, and so Borel. Since this sequence of sets is descending (with respect to m, for given k and i) and its intersection is empty, for every $l = 1, 2, ...$ there exists an m such that

$$\zeta(B^i_k(m)) \leq \frac{1}{2^{k+i+l}};$$

denote by $m^i_k(l)$ any such m.

Let now

$$A^i_k(l) \equiv \Omega - C(s^i_m) \text{ for } m = m^i_k(l),$$

$$A_k(l) \equiv \bigcap_{i=1}^{\infty} A^i_k(l),$$

and

$$\hat{A}(l) \equiv \bigcup_{k=1}^{\infty} A_k(l) \cup \bigcup_{i=1}^{\infty} \{s^i\}.$$

The set $A(l)$ is a countable union of closed sets with empty interior by an argument analogous to that used in the case of the sets $\hat{A}_P$, and thus so is

$$\hat{A} \equiv \bigcup_{l=1}^{\infty} \hat{A}(l).$$

To show that (8.6)-(8.7) are satisfied, notice that, by the definition of $B_k^i(m)$ for $m = m_k^i(l)$, if $P \notin B_k^i(m)$, then $C(s_m^i) \subset C_P^k(s_t^i)$ for $t = t(i,k,P)$; therefore

$$\text{if } P \in \Delta(\Omega) - \bigcup_{i=1}^{\infty} B_k^i(m_k^i(l)), \text{ then } A_k(P) \subset A_k(l),$$

which in turn yields that

$$\text{if } P \in \Delta(\Omega) - \bigcup_{k=1}^{\infty}\bigcup_{i=1}^{\infty} B_k^i(m_k^i(l)), \text{ then } \hat{A}_P \subset \hat{A}(l).$$

Thus,

$$B \equiv \bigcup_{l=1}^{\infty} \left[\Delta(\Omega) - \bigcup_{k=1}^{\infty}\bigcup_{i=1}^{\infty} B_k^i(m_k^i(l)) \right] \subset \{P \in \Delta(\Omega) : \hat{A}_P \subset \hat{A}\}.$$

It remains to show that $\zeta\,(B) = 1$; however

$$\zeta \left(\Delta(\Omega) - \bigcup_{k=1}^{\infty}\bigcup_{i=1}^{\infty} B_k^i(m_k^i(l)) \right) \geq 1 - \sum_{k=1}^{\infty}\sum_{i=1}^{\infty} \zeta\left(B_k^i(m_k^i(l))\right) \geq$$

$$\geq 1 - \sum_{k=1}^{\infty}\sum_{i=1}^{\infty} \frac{1}{2^{k+i+l}} = 1 - \frac{1}{2^l}$$

for every $l \in N$.■

Lemma 3. *For every $t \in N$, the set*

$$B_k^i(t) \equiv \{P \in \Delta(\Omega) : t(i,k,P) > t\}$$

is open.

Proof: Let $P \in B_k^i(t)$. By definition, $t(i,k,P) > t$, which means that

$$P(C(s_t^i) - \{s^i\}) > \frac{1}{2^{k+i}}.$$

Further, there exists an $m > t$ such that

$$P(C(s_t^i) - C(s_m^i)) > \frac{1}{2^{k+i}}; \tag{8.8}$$

indeed, the sequence sets $C(s_t^i) - C(s_m^i)$ is ascending (as m goes to infinity), and its union is equal to $C(s_t^i) - \{s^i\}$.

Note that each cylinder is an open and closed subset of Ω, and so is $C(s_t^i) - C(s_m^i)$. Thus, the function $f : \Omega \to R$ given by:

$$\begin{aligned} f(s) &= 1 \text{ for every } s \in C(s_t^i) - C(s_m^i) \\ f(s) &= 0 \text{ for every } s \notin C(s_t^i) - C(s_m^i) \end{aligned}$$

is continuous.

Let

$$\delta \equiv P(C(s_t^i) - C(s_m^i)) - \frac{1}{2^{k+i}}, \tag{8.9}$$

and let the set $N(P)$ stands for all measures $Q \in \Delta(\Omega)$ such that

$$\left| \int f dQ - \int f dP \right| < \delta.$$

This last inequality means that

$$\left| Q(C(s_t^i) - C(s_m^i)) - P(C(s_t^i) - C(s_m^i)) \right| < \delta;$$

by (8.8) and (8.9),

$$Q(C(s_t^i) - \{s^i\}) \geq Q(C(s_t^i) - C(s_m^i)) > 1/2^{k+i},$$

which implies that $Q \in B_i^k(t)$.

That is, the set $N(P)$, which is an open neighborhood of P in weak*-topology, is contained in $B_k^i(t)$.■

Corollary 2. *For every $s \in \Omega$, the set $\{P \in \Delta(\Omega) : s \in \hat{A}_P\}$ is Borel.*

Proof: If $s = s^i$ for some $i \in N$, then $\{P \in \Delta(\Omega) : s \in \hat{A}_P\} = \Delta(\Omega)$ by definition. Suppose therefore that $s \neq s^i$ for any $i \in N$. It suffices to show that the sets $\{P \in \Delta(\Omega) : s \in C_P^k(s_t^i), t = t(i,k,P)\}$ are Borel, as

$$\{P \in \Delta(\Omega) : s \in \hat{A}_P\} = \Delta(\Omega) - \left[\bigcap_{k=1}^{\infty}\left(\bigcup_{i=1}^{\infty}\{P \in \Delta(\Omega) : s \in C_P^k(s_t^i)\}\right)\right].$$

Take $m = 0, 1, ...$ such that $s \in C_P^k(s_m^i) - C_P^k(s_{m+1}^i)$; since $s \neq s^i$ and $C_P^k(s_0^i) = \Delta(\Omega)$, such m is unique. Observe now that

$$\{P \in \Delta(\Omega) : s \in C_P^k(s_t^i), t = t(i,k,P)\} = \Delta(\Omega) - \{P \in \Delta(\Omega) : t(i,k,P) > m\},$$

and so the set $\{P \in \Delta(\Omega) : s \in C_P^k(s_t^i), t = t(i,k,P)\}$ is closed by lemma 3.

Proof of Proposition 7: We prove proposition 7 under a weaker assumption that the test T does not reject an informed expert with probability $1-\varepsilon$. Notice first that proposition 2 also applies (by a similar argument) to tests that do not reject an informed expert with probability $1-\varepsilon$. That is, there exists a rejection test T^δ that is harder then T and does not reject an informed expert with probability $1-\varepsilon-\delta/2$. By construction, the rejection sets $(A_P)^c$ (of the test T^δ) are a countable union of disjoint cylinders C_P^i, $i \in N$, with base on finite histories. Let $T^{\delta,m}$ be the test such that the rejection sets $(A_P^m)^c$ are the finite union of all cylinders C_P^i with base on a history on $\{0,1\}^m$. By definition, T^δ is harder than $T^{\delta,m}$. So, $T^{\delta,m}$ does not reject an informed expert with probability $1-\varepsilon-\delta/2$. Moreover, $T^{\delta,m}$ is a finite test. Hence, by proposition 1, there exists $\zeta_{T,m}$ such that for every history $s \in \{0,1\}^\infty$

$$\zeta_{T,m}\left\{P \in \Delta(\Omega) \mid T^{\delta,m}(s,P) = 0\right\} \leq \varepsilon + \delta.$$

Let s_m be any finite history on $\{0,1\}^m$ such that $T(s_m,P) = 0$. Given that T^δ is harder than T it follows that $T^\delta(s_m,P) = 0$. By definition, $C(s_m) \subseteq (A_P)^c$ and, therefore, $C(s_m) \subseteq (A_P^m)^c$. So, $T(s_m,P) = 0$ implies that $T^{\delta,m}(s_m,P) = 0$. Thus,

$$\{P \in \Delta(\Omega) \mid T(s_m,P) = 0\} \subseteq \left\{P \in \Delta(\Omega) \mid T^{\delta,m}(s_m,P) = 0\right\}.$$

So, for every $s_m \in \{0,1\}^m$, $\zeta_{T,m}\{P \in \Delta(\Omega) \mid T(s_m, P) = 0\} \leq \varepsilon + \delta$.■

Proof of Proposition 8. Fix $\varepsilon > 0$, $\delta > 0$ and $K \in N$. We first show that there exist $b > 0$ and $\bar{m} \in N$ such that if $m \geq \bar{m}$ then T^{β} does not reject any theory $P \in \Lambda_{m,\delta}$ with probability $1 - \varepsilon$.

Step 1: For every $\delta > 0$, there exists $b > 0$ such that

$$\left(\frac{1}{2} + b\right)^{\left(\frac{1}{2}+\frac{\delta}{4}\right)} \cdot \left(\frac{1}{2} - b\right)^{\left(\frac{1}{2}-\frac{\delta}{4}\right)} > \frac{1}{2}. \tag{8.10}$$

Indeed, notice that at $b = 0$ the left-hand side is equal to the right-hand side and the derivative of the right-hand side (at $b = 0$) is equal to $\delta/2 > 0$.

Fix such a $b > 0$.

Step 2: Consider the Bernoulli scheme with the probability of success equal to $\frac{1}{2} + \frac{\delta}{2}$. By the law of large numbers, there exists an $\overline{n} \in N$ such that for every $n \geq \overline{n}$ the success rate exceeds $\frac{1}{2} + \frac{\delta}{4}$ with probability $1 - \varepsilon$ (or higher). By step 1, we can additionally assume that

$$\left(\frac{\left(\frac{1}{2} + b\right)^{\left(\frac{1}{2}+\frac{\delta}{4}\right)} \cdot \left(\frac{1}{2} - b\right)^{\left(\frac{1}{2}-\frac{\delta}{4}\right)}}{\frac{1}{2}}\right)^{\overline{n}} \geq K. \tag{8.11}$$

Given any $P \in \Lambda_{m,\delta}$ and $s \in \Omega$, denote by $n_m(P, s)$ the number of periods $t \leq m$ such that

$$f_{t-1}^P(s) \in [0, \frac{1}{2} - \frac{\delta}{2}) \text{ or } f_{t-1}^P(s) \in (\frac{1}{2} + \frac{\delta}{2}, 1].$$

There exists $\bar{m} \in N$ such that if $m \geq \bar{m}$ then for any $P \in \Lambda_{m,\delta}$ and $s \in \Omega$ that number $n_m(P, s)$ must exceed $\overline{n}$.

Indeed, if $f_{t-1}^P(s) \in [\frac{1}{2} - \frac{\delta}{2}, \frac{1}{2} + \frac{\delta}{2}]$ for all but (at most) $\overline{n}$ of the periods $t \leq m$, then

$$\frac{1}{m}\sum_{t=1}^{m}(f_{t-1}^P(s) - 0.5)^2 \leq \frac{1}{m}\left[\overline{n}(0.5)^2 + (m - \overline{n})\left(\frac{\delta}{2}\right)^2\right];$$

since the right-hand side of this inequality cannot exceed δ for sufficiently large m, we obtain a contradiction with $P \in \Lambda_{m,\delta}$.[23]

Let now $\widehat{n}_m(P,s)$ denote the number of periods $t \leq m$ that satisfy the following condition: $f_{t-1}^P(s) \in [0, \frac{1}{2} - \frac{\delta}{2})$ and $I_t(s) = 0$ or $f_{t-1}^P(s) \in (\frac{1}{2} + \frac{\delta}{2}, 1]$ and $I_t(s) = 1$. By construction, for any $m \geq \bar{m}$ and $P \in \Lambda_{m,\delta}$ the set

$$B(m,P) = \left\{ s \in \Omega : \frac{\widehat{n}_m(P,s)}{n_m(P,s)} > \frac{1}{2} + \frac{\delta}{4} \right\}$$

has $P-$probability $1 - \varepsilon$ (or larger). Thus, if $s_m = s \mid m$ and $s \in B(m,P)$ then

$$\begin{aligned} \bar{P}(C(s_m)) &\geq \left(\frac{1}{2} + b\right)^{\left(\frac{1}{2}+\frac{\delta}{4}\right)n} \cdot \left(\frac{1}{2} - b\right)^{\left(\frac{1}{2}-\frac{\delta}{4}\right)n} \cdot \left(\frac{1}{2}\right)^{m-n} = \\ &= \left(\frac{\left(\frac{1}{2} + b\right)^{\left(\frac{1}{2}+\frac{\delta}{4}\right)} \cdot \left(\frac{1}{2} - b\right)^{\left(\frac{1}{2}-\frac{\delta}{4}\right)}}{\frac{1}{2}} \right)^n \cdot \left(\frac{1}{2}\right)^m, \end{aligned}$$

where $n = n_m(P,s)$. By 8.11, if $m \geq \bar{m}$ then $n_m(P,s) \geq \overline{n}$ and, by construction, $B(m,P)$ has $P-$probability $1 - \varepsilon$ (or larger). So, if $m \geq \bar{m}$ then $\bar{P}(C(s_m))$ exceeds $K(0.5)^m$ with $P-$probability $1 - \varepsilon$ (or larger).

We now show that if $K\varepsilon > 1$ then rejection by the test T^β cannot be delayed, with probability ε. Let $I \equiv \{1, ..., 2^m\}$. Denote by $F \equiv \left\lceil \frac{2^m}{K} \right\rceil$ the largest integer no larger than $\frac{2^m}{K}$. Let Z be the set of all $z \in I^F$ such that no two coordinates of z are equal. By Z_i, $i \in I$, denote the subset of Z that consists all elements $z \in Z$ such that one (and only one) coordinate of z is equal to i. Let $x_z \in [0,1]$, $z \in Z$.

Step 1: We shall show first that if

$$\sum_{z \in Z_i} x_z \geq \varepsilon, \tag{8.12}$$

for all $i \in I$, then

$$\sum_{z \in Z} x_z \geq K\varepsilon. \tag{8.13}$$

[23] We assume here that $\delta < 4$, but if $\delta \geq 4$, then $\Lambda_{m,\delta} = \emptyset$.

Indeed, any $z \in Z$ belongs to a set Z_i if (and only if) i is one of the coordinates of z; so, it belongs to exactly F such sets. Thus,

$$F \sum_{z \in Z} x_z = \sum_{i \in I} \sum_{z \in Z_i} x_z \geq 2^m \varepsilon.$$

Step 2: A theory P passes the test T^β only on the histories s_m to which $\bar{P}$ assigns probability greater than $\frac{K}{2^m}$. For each P, there are at most F histories s_m with this property. Let T' be any test easier than T such that each theory P passes T' on exactly F histories s_m.

Identify I with the set of all possible histories s_m. Given a $z \in Z$, let Λ_z stand for the set of all probability measures that pass T' on all histories s_m from the range of z. Given $z \neq z'$, $\Lambda_z \cap \Lambda_{z'} = \varnothing$ because no theory can pass T' on the ranges of z and z' simultaneously (together they contain more than F histories s_m). Fix $\zeta \in \Delta\Delta(\Omega)$. Let $x_z \equiv \zeta(\Lambda_z)$. Assume, to the contrary, that ζ can delay rejection of T' by m periods with probability ε. Then, by definition, (8.12) is satisfied and, by Step 1, so is (8.13). It is, however, a contradiction because

$$1 \geq \sum \zeta(\Lambda_z) \geq K\varepsilon > 1.$$

■

Proof of Proposition 9: (**i**) Take any probability measure $P \in \Lambda_{m,\delta}$. For any $s \in \Omega$ and $t \leq m$ define functions $g_{s,t}, h_{s,t} : \Omega \to R$ by

$$g_{s,t}(r) = 1 \text{ if } r_{t-1} = s_{t-1} \text{ and } g_{s,t}(r) = 0 \text{ otherwise,}$$

$$h_{s,t}(r) = 1 \text{ if } r_t = s_t \text{ and } h_{s,t}(r) = 0 \text{ otherwise.}$$

By taking ε small enough, we guarantee that if

$$\forall_{s \in \Omega} \quad \left| E^P g_{s,t} - E^Q g_{s,t} \right| < \varepsilon,$$

then the measure assigned by P and Q to the cylinder $C(s_{t-1})$ are as close as we wish; by taking ε small enough, we also guarantee that if

$$\forall_{s \in \Omega} \quad \left| E^P h_{s,t} - E^Q h_{s,t} \right| < \varepsilon,$$

then the measure assigned by P and Q to the cylinder $C(s_t)$ are as close as we wish. Thus if a probability measure Q satisfies the two inequalities for any $s \in \Omega$ and $t \leq m$, then $f^P_{t-1}(s)$ and $f^Q_{t-1}(s)$ are as close as we wish, which (by definition) implies that $Q \in \Lambda_{m,\delta}$, as so does P.

(**ii**) Take any $P \in \Delta(\Omega)$, continuous functions $h_1, ..., h_l : \Omega \to R$, and positive real numbers $\varepsilon_1, ..., \varepsilon_l$. It follows from the continuity of $h_1, ..., h_l$ and the compactness of Ω that there exists a (large enough) $k \in N$ such that

$$r_k = s_k \Longrightarrow \forall_{i=1,...,l} \quad |h_i(r) - h_i(s)| < \varepsilon_i.$$

Thus, if two probability measures P and Q have the property that

$$\forall_{s\in\Omega} \; P(C(s_k)) = Q((s_k)), \tag{8.14}$$

then

$$\forall_{i=1,...,l} \quad \left|E^P h_i - E^Q h_i\right| < \varepsilon_i.$$

Take a probability measure Q satisfying (8.14) with the following property: for any $s \in \Omega$ and $t > k$

$$f^Q_k(t) = q \text{ for some } q \in (\overline{\delta}, 1], \tag{8.15}$$

where

$$\left(\overline{\delta} - 0.5\right)^2 = \delta.$$

By (8.15), $Q \in \Lambda_{m,\delta}$ whenever m is sufficiently large, and by (8.14), Q belongs to the neighborhood of the probability measure P determined by $h_1, ..., h_l : \Omega \to R$ and $\varepsilon_1, ..., \varepsilon_l$.■

References

[1] Anderson, R. and W. Zame (2001) "Genericity with Infinitely Many Parameters," *Advances in Theoretical Economics* **1**, 1-62.

[2] Brier, G. (1950), "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review* **78**, 1-3.

[3] Cahn, A. (2005), "General Procedures Leading to Correlated Equilibria," forthcoming *International Journal of Game Theory*.

[4] Dawid, A. (1982), "The Well-Calibrated Bayesian," *Journal of the American Statistical Association* **77**, 605-613.

[5] Dawid, A. (1985), "Calibration-Based Empirical Probability," *The Annals of Statistics* **13**, 1251-1274.

[6] Dekel, E. and Y. Feinberg (2006), "Non-Bayesian Testing of a Stochastic Prediction," forthcoming *Review of Economic Studies*.

[7] Dudley, R.M. (1989), *Real Analysis and Probability*, Wadsworth Inc., Belmont, California.

[8] Engelking, R. (1989), *General Topology*, Heldermann Verlag, Berlin.

[9] Fan, K. (1953), "Minimax Theorems," *Proceedings of the National Academy of Science U.S.A.*, **39**, 42-47.

[10] Foster, D. and R. Vohra (1998), "Asymptotic Calibration," *Biometrika* **85**, 379-390.

[11] Fudenberg, D. and D. Levine (1995), "Consistency and Cautious Fictitious Play," *Journal of Economic Dynamics and Control*, **19**, 1065-1089.

[12] Fudenberg, D. and D. Levine (1999a) "Conditional Universal Consistency," *Games and Economic Behavior* **29**, 104-130.

[13] Fudenberg, D. and D. Levine (1999b) "An Easier Way to Calibrate," *Games and Economic Behavior*, **29**, 131-137.

[14] Hart, S. (2005) "Adaptative Heuristics" *Econometrica*, **73**, 1401 - 1430

[15] Hart, S. and A. Mas-Colell (2000) "A Simple Adaptive Procedure Leading to Correlated Equilibrium", *Econometrica*, **68**, 1127-1150.

[16] Hart, S. and A. Mas-Colell (2001) "A General Class of Adaptative Strategies," *Journal of Economic Theory*, **98**, 26-54.

[17] Hayek, F. (1945) "The Use of Knowledge in Society," *American Economic Review*, **35**, 519-530.

[18] Kalai, E., E. Lehrer, and R. Smorodinsky (1999) "Calibrated Forecasting and Merging," *Games and Economic Behavior* **29**, 151-169.

[19] Lehrer, E. (2001) "Any Inspection Rule is Manipulable," *Econometrica* **69**, 1333-1347.

[20] Lehrer, E. and E. Solan (2003), "No Regret with Bounded Computation Capacity," Tel Aviv University, mimeo.

[21] Oxtoby, J. (1980), *Measure and category: A survey of the analogies between topological and measure spaces*, Springer Verlag, New York-Berlin.

[22] Popper, K. (1968), *The Logic of Scientific Discovery*, Hutchinson & Co., London.

[23] Rudin, W. (1973), *Functional Analysis*, McGraw-Hill, Inc.

[24] Rustichini, A. (1999), "Optimal Properties of Stimulus-Response Learning Models," *Games and Economic Behavior*, **29**, 244-273.

[25] Sandroni, A., R. Smorodinsky and R. Vohra (2003), "Calibration with Many Checking Rules," *Mathematics of Operations Research* **28**, *141-153.*

[26] Seidenfeld, T. (1985), "Calibration, Coherence and Scoring Rules," *Philosophy of Science*, **52** 274-294.

[27] Shiryaev, A. (1996), *Probability*, Springer Verlag, New York Inc.

[28] Stinchcombe, M. (2001) "The Gap Between Probability and Prevalence: Loneliness in Vector Spaces," *Proceedings of the American Mathematical Society*, **129**, 451-457.

[29] Vovk, V. and G. Shafer (2004), "Good Randomized Sequential Probability Forecasting is Always Possible," mimeo, University of London.