No Manipulation Results for Non-Bayesian Tests

Eddie Dekel* and Yossi Feinberg†

Preliminary, July 2005‡

Abstract

In Dekel and Feinberg (2004) we suggested a test for discovering whether a potential expert is informed of the distribution of a stochastic process. This *category test* requires predicting a "small"– category I – set of outcomes. In this paper we show that there is a randomized category test that cannot be manipulated, i.e. such that no matter how the potential expert randomizes his prediction, there will be realizations where he will fail to pass the test with probability 1. The set of outcomes where he fails can be made large – a category II set – under the continuum hypothesis. Moreover, these results hold for the finite approximations of the category tests where the non-expert is failed in finite time and the expert is failed with small probability. **JEL Classification:** K9

1 Introduction

In Dekel and Feinberg (2004) we suggested a test to determine whether a potential expert knows the distribution governing a stochastic process. The tester is completely uninformed and non-Bayesian, in the sense that she does not have a prior distribution over the possible distributions that govern the stochastic process, nor does she have a prior over the probability that she is facing an expert. We showed that for each predicted probability measure μ there exists a (category I) set S_μ such that $\mu(S_\mu) = 1$ and such that the set of measures that assign positive probability to S_μ is small, in the sense that it is a category I set of measures in the space of probability measures.[1] Thus "most" predictions other than those made by

* *Department of Economics, Northwestern University, Evanston, USA and Tel-Aviv University, Tel-Aviv, Israel.* e-mail: dekel@northwestern.edu

† *Stanford Graduate School of Business, Stanford CA 94305-5015.* e-mail: yossi@gsb.stanford.edu

‡ Please contact authors before quoting.

[1] A category I set of outcomes is a countable union of nowhere dense sets—sets whose closure has an empty interior.

an expert who knows the actual distribution will almost surely fail, and the knowledgeable expert will almost surely pass. We also provided a finite approximation for this test: for any given $\varepsilon > 0$ any prediction can be tested with a set that will fail all but a category I set of predictions in finite time and will fail the expert with probability of no more than ε.

However, an uninformed expert might still be able to make a randomized prediction that would pass a test with high probability (with respect to his randomized strategy), i.e., one could potentially *manipulate* the test. In this paper we provide a randomized category test that cannot be manipulated. The randomized test picks a test from the class of tests developed in our previous work according to a distribution that is known to the potential expert. However, no matter how an uninformed expert chooses his prediction he will fail the test on a set of realizations the size of the continuum. To address the concern that this set—on which a potential manipulator will fail—might nevertheless be small, we show that (assuming the continuum hypothesis) we can assure that the failure will occur on a set larger than a category I set, i.e. a set that cannot be covered by a countable union of nowhere dense sets. This failure will occur with probability 1 with respect to the randomized category test employed. It turns out that the finite approximation of the randomized test also is not manipulable: the uninformed expert is assured to fail on a continuum even though the test must determine the non-expert in finite time while not failing the expert with high probability.

These results distinguish such "category" tests from the well studied calibration tests not only in the formal interpretation of what a test is supposed to accomplish, but also in whether a non-expert can manipulate the test.[2] As was shown by Lehrer (2001) a randomized calibration test can be passed by a completely uninformed non-expert. Moreover, Kalai, Lehrer and Smorodinsky (1999), Fudenberg and Levine (1999), Sandroni, Smorodinsky and Vohra (2003), and Sandroni (2003) provide generalizations and variations of such results to a large class of tests which raised the question whether an unmanipulable test could actually exist. In these papers a randomized strategy by the non-expert can pass a randomized calibration test on *every* realization with probability 1 (with respect to the random prediction). In contrast, our randomized category test guarantees failure on a set of outcomes that is not small no matter what randomized strategy the non-expert uses (obviously without hindering the guarantee that the informed expert will pass the test).

The conceptual idea behind the construction of an unmanipulable random category test is based on the observation that there are deterministic category tests that can only be passed on a large set of realizations if the prediction made (whether deterministic or random) has a specific property. By randomizing over such category tests we can assure that these specific

[2]See Dawid (1982,1985) and Foster and Vohra (1998) for early papers on calibration.

restrictions cannot be simultaneously satisfied for a collection of tests that are assigned positive probability. Hence, we build on the unique nature of category tests insofar that they can be quite demanding in the predictions required from an expert.

2 Testing and manipulation: definitions

Consider the set of realizations of a stochastic process $\Omega = \{0,1\}^{\aleph_0}$ governed by a probability distribution. Let $\Delta(\Omega)$ denote the set of probability measures over Ω endowed with the σ-field generated by the finite cylinders. Let $t : \Delta(\Omega) \longrightarrow 2^{\Omega}$ denote a test. The interpretation is that if a predictor proposes the distribution P then he passes the test t if and only if the realization $\omega \in \Omega$ of the process satisfies $\omega \in t(P)$.

Calibration tests have been shown to be susceptible to manipulation, in the sense that if t is a calibration test which satisfies

$$P(t(P)) = 1 \text{ for every } P \in \Delta(\Omega) \tag{1}$$

then there exists a randomized prediction $\nu \in \Delta(\Delta(\Omega))$ such that, for every $\omega \in \Omega$, $\nu\{Q|\omega \in t(Q)\} = 1$. In other words, by randomly predicting Q an uninformed predictor will pass the calibration test t at every realization of the process.[3] Furthermore, even if the tester chooses a calibration test t at random according to some given probability distribution over the collection of calibration tests, there exists a randomized prediction ν that passes the test with probability one with respect to the randomly selected test at every realization ω and the randomly selected prediction according to ν. In other words, calibration tests can be manipulated.

Clearly the property in (1) is required of a test if we do not wish to rule out a potential expert who is actually informed as to the distribution governing the process. In Dekel and Feinberg (2004) we defined a new class of tests that we now call *category tests*. This class is defined as:

$$T_C = \{\text{all tests } t \text{ such that for every } P \text{ we have } t(P) \text{ is category I set and } P(t(P)) = 1\} \tag{2}$$

[3] A typical calibration test in the literature does not apply to an ex ante distribution such as P, but instead asks for predictions conditional on the history as a realization unveils. The empirical distribution of the true state ω is then compared to the predictions. Of course the collection of all such conditionals determines P if one also considers all realizations. A randomized prediction (cf. Foster and Vohra(1998)) randomizes the conditional predictions and corresponds to a randomization ν over the space of measures $\Delta(\Omega)$. Therefore, the manipulation results can be applied to ex ante predictions as well.

By definition category tests satisfy the property in (1). The definition in (2) has the additional property that for every test $t \in T_C$ for any prediction Q the predictor will fail the test on a *large* set of realizations—the complement of a category I set which is therefore a category II set of realizations (note that the complement of a category II set need not be a category I set). In Dekel and Feinberg (2004) we showed that the class of tests T_C is non-empty and most importantly that the set of measures that assign positive probability to a category I set is itself a category I set in the space of measures. Hence, it seems difficult to pass our category tests without knowing the actual distribution governing the process.

Nevertheless, the question arises whether a non-expert can manipulate tests by using a randomized prediction, i.e. choosing at random a prediction from $\Delta(\Omega)$. We now define these notions formally. We say that a predictor can *manipulate* a test t if there exist a probability distribution $\mu \in \Delta(\Delta(\Omega))$ such that for all $\omega \in \Omega$, with μ probability 1 we have $\omega \in t(Q)$. Formally,

$$\mu(\{Q|\omega \in t(Q)\}) = 1, \text{ for all } \omega \in \Omega. \tag{3}$$

A *random test* is a distribution $\lambda \in \Delta(T_C)$, where once a prediction P is given the test t is chosen according to λ and applied to the realization.[4] A random test λ can be manipulated for sure if there exists a randomized prediction $\mu \in \Delta(\Delta(\Omega))$ such that

$$\lambda(\{t \in T_C|\mu(\{Q|\omega \in t(Q)\}) = 1 \text{ , for all } \omega \in \Omega\}) = 1. \tag{4}$$

That is, almost all predictions pass almost all tests for every realization. This is a very strong sense in which manipulation occurs, and as shown by Lehrer (2001), if instead of T_C we considered calibration tests then all randomized calibration tests would be manipulable in this sense.

If we were to show that the above cannot hold for category tests, then we would be showing that there is a randomized category test that fails every randomized prediction with positive probability for some realization. This is a start towards showing that our tests are not manipulable, but it is a weak result. Instead we show a stronger result: there exist randomized category tests that fail all randomized predictions *with probability* **1** (according to the random measure used for the test and prediction) *for "many" realizations*. Naturally, the largest this set of realizations could be is a complement of a category I set, since the true prediction passes on a category I set of realizations. We have only been able to obtain the following weaker but still strong result: there is a randomized test that fails all randomized predictions with probability 1 on an uncountable set of realizations. Moreover, under the

[4] More preciesly, a random (category) test is given by a σ-field on T_C and a measure on the resulting $\Delta(T_C)$.

continuum hypothesis, the set can be made a category II set—a set that cannot be covered by a countable union of nowhere dense sets.

Formally, there exists a σ-field on T_C and a $\lambda \in \Delta(T_C)$ such that for all $\mu \in \Delta(\Delta(\Omega))$

$$\lambda(\{t \in T_C | \exists \text{ an uncountable set } S \subset \Omega \text{ such that } \mu(\{Q | S \cap t(Q) = \emptyset\}) = 1\}) = 1. \quad (5)$$

Hence, not only does the failure to manipulate occur with probability 1, it also occurs on an uncountable set of realizations. Under the continuum hypothesis, there is a randomized test such that each of the sets S considered in (5) can be chosen to be category II sets. As before, the set S may vary with the randomly selected test t.

3 A preliminary result

Given $\mu \in \Delta(\Delta(\Omega))$ we define the measure $\bar{\mu} \in \Delta(\Omega)$ as:

$$\bar{\mu}(E) = \int_{\Delta(\Omega)} P(E) d\mu(P) \quad (6)$$

for every measurable set E. This measure is sometimes refereed to as the "center of gravity" of the measure μ. Note that since Ω is a compact metric space so is $\Delta(\Omega)$ in the weak* topology (cf. Theorem 6.4 in Parthasarathy (1967)). By the definition of the weak* topology we have that for every continuous function $f \in C(\Omega)$ the functional $f(P) = \int_\Omega f(\omega) dP(\omega)$ is a continuous functional on $\Delta(\Omega)$. In particular the continuous functionals on $\Delta(\Omega)$ separate points. From the convexity and compactness of $\Delta(\Omega)$ in the weak* topology we have that the generalized integral $\int_{\Delta(\Omega)} P d\mu(P)$ exists in the sense that for every linear functional Λ on $\Delta(\Omega)$ we have

$$\Lambda(\bar{\mu}) = \int_{\Delta(\Omega)} (\Lambda(P)) d\mu(P) \quad (7)$$

and $\bar{\mu}$ is a probability measure. See Theorems 3.27 and 3.28 in Rudin (1991). Since the measure $\bar{\mu}$ must satisfy

$$\int_\Omega f(\omega) d\bar{\mu} = \int_{\Delta(\Omega)} (\int_\Omega f(\omega) dP) d\mu(P) \quad (8)$$

for every continuous function f we have that regularity implies that (6) is well defined. To see this, consider first a closed set E, we have that $\nu(E) = \inf\{\int f d\nu | f \geq \chi_E\}$ where χ_E is the characteristic function of E. In particular, this holds for $\nu = \bar{\mu}$ as well. By regularity $\nu(G) = \sup\{\nu(E)|E \text{ is closed, } E \subset G\}$ for every measurable set G. Hence we have measurability of $P(E)$ for measurable sets E and $\int_{\Delta(\Omega)} P(E)d\mu(P)$ is defined and coincides with $\bar{\mu}$ as required.

The following Lemma makes use of this center of gravity measure.

Lemma 1 *For every measure $\mu \in \Delta(\Delta(\Omega))$ there exists a test $t \in T_C$ and a category II set S such that $\mu(\{P \in \Delta(\Omega)|S \bigcap t(P) = \emptyset\}) = 1$.*

Hence the measure μ leads to failure of the test over a category II set of points. (This is not the result we seek as the test depends on μ, but it is an instructive first step.)

Proof. Given μ let $\bar{\mu} \in \Delta(\Omega)$ be defined as in (6).

Let S be a category II set such that $\bar{\mu}(S) = 0$. Such a set exists since for every measure over Ω we can find a category I set that has measure 1 and pick S as its complement. Consider the following test:

$$t_S(P) = \begin{cases} \text{any category I set } R \text{ such that } P(R) = 1 & \text{if } P(S) > 0 \\ \text{a category I set } R \text{ s.t. } R \bigcap S = \emptyset \text{ and } P(R) = 1 & \text{if } P(S) = 0 \end{cases} \tag{9}$$

The test is well defined since by taking Q to be a category I set such that $P(Q) = 1$ (which exists as noted above), then if $P(S) = 0$ one can set $R = Q \setminus S$ and have the required category I set for the test.

Since $\bar{\mu}(S) = 0$ we have $\int_{\Delta(\Omega)} P(S)d\mu(P) = 0$ and so we must have $\mu(\{P|P(S) > 0\}) = 0$ so $\mu(\{P|P(S) = 0\}) = 1$ but every P with $P(S) = 0$ fails the test t_S at every $\omega \in S$ as required. ∎

The idea behind the test constructed in (9) is that the test can force failure on a given category II set – S – whenever a predicted measure assigns zero probability to the set S.

4 Unmanipulable random category tests

Proposition 2 *There exists a randomized test λ such that for every $\mu \in \Delta(\Delta(\Omega))$ with μ-probability 1 the non-expert fails on an uncountable set of points (with λ-probability 1 with respect to the distribution over the tests): $\lambda(\{t \in T_C|\exists \text{ an uncountable set } S \subset \Omega \text{ such that } \mu(\{Q|S \cap t(Q) = \emptyset\}) = 1\}) = 1$.*

Proof. Fix an arbitrary category test t. For every real number $r \in [0, 1]$ define the sets

$$F_r = \{\omega \in \Omega| \text{ the average of 1's in } \omega \text{ converges to } r\}. \tag{10}$$

For every r consider a category test t_r as follows:

$$t_r(P) = \begin{cases} t(P) & \text{if } P(F_r) > 0 \\ t(P) \setminus F_r & \text{if } P(F_r) = 0 \end{cases} \tag{11}$$

The tests t_r are category tests since any subset of a category I set is category I and hence $t(P) \setminus F_r$ is a category I set. These test also satisfy $P(t_r(P)) = 1$ since t is a category test and since

$$P(t_r(P)) = P(t(P) \setminus F_r) = \begin{cases} P(t(P)) & \text{if } P(F_r) > 0 \\ P(t(P) \setminus F_r) = P(t(P)) & \text{if } P(F_r) = 0 \end{cases} \tag{12}$$

for all P.

Denote by $f : [0,1] \longrightarrow \{t_r\}_{r\in[0,1]}$ the map associating with each index $r \in [0,1]$ a category test as above. Assume first that the tests $\{t_r\}_{r\in[0,1]}$ are disjoint, i.e. for all $r \neq s$ we have $t_r \neq t_s$. Let λ be any non-atomic probability measure over the set of tests $\{t_r\}_{r\in[0,1]}$ as a measure space endowed from the Borel measurable sets on the unit interval, e.g. a uniform distribution picking the index of the test $r \in [0,1]$. Consider any given distribution $\mu \in \Delta(\Delta(\Omega))$ and $\bar{\mu}$ as defined above in (6). The random prediction μ will pass a test t_r with positive probability on all but a countable set of ω's only if $\bar{\mu}(F_r) > 0$ since if $\bar{\mu}(F_r) = 0$ then with μ-probability 1 the measure P selected will satisfy $P(F_r) = 0$ and will fail on the uncountable set F_r. Since $\bar{\mu}$ is a probability measure we have that there are at most a countable number of the disjoint sets F_r to which $\bar{\mu}$ can assign positive probability. But since λ is non-atomic it assigns zero probability to any countable set of indices. In particular with λ-probability 1 an index r is chosen such that $\bar{\mu}(F_r) = 0$ and with μ-probability 1 the predictor fails on an uncountable set of points.

Now consider the case where the set of tests $\{t_r\}_{r\in[0,1]}$ is not disjoint but every test appears in $\{t_r\}_{r\in[0,1]}$ at most a countable number of times. The mapping $f : [0,1] \longrightarrow \{t_r\}_{r\in[0,1]}$ induces a measurable space on $\{t_r\}_{r\in[0,1]}$ by defining $S \subset \{t_r\}_{r\in[0,1]}$ to be measurable if and only if $f^{-1}(S)$ is measurable. Since $f^{-1}(\{t_r\})$ is countable we have that every singleton is measurable. Any non-atomic probability measure λ on $[0,1]$ induces a non-atomic probability measure λ_f on $\{t_r\}_{r\in[0,1]}$ by defining $\lambda_f(S) = \lambda(f^{-1}(S))$ for every measurable set $S \subset \{t_r\}_{r\in[0,1]}$. As in the previous case, for any measure $\mu \in \Delta(\Delta(\Omega))$ at most a countable collection of tests can be passed simultaneously on all but a countable set of realizations. Denote this collection of tests by $\bar{S} \subset \{t_r\}_{r\in[0,1]}$. Clearly $\bar{S}$ is measurable and $\lambda_f(\bar{S}) = 0$.

Finally, if there is a category test $\bar{t}$ such that $\bar{t} = t_r$ for an uncountable collection of indices r, then we simply consider the deterministic tests $\bar{t}$. A randomized prediction μ

cannot pass the deterministic test $\bar{t}$ at all but a countable set of realizations since that will require that μ have an atom in each of the uncountable collection of indices for which $\bar{t} = t_r$. In this case the deterministic category test $\bar{t}$ cannot be manipulated.

Remark 3 *The tests $\{t_r\}_{r\in[0,1]}$ are constructed, so one might (even though we do not) know which case above they fall into: disjoint, countably many repetitions, or uncountably many repetitions. If we knew the latter case held, then—as noted—we have the much stronger result showing a deterministic test that cannot be manipulated on an uncountable set of realizations.*

■

A set with an uncountable number of points can still be "small." For instance, it might be a category I set. Our next result assures that a non-expert will fail with probability 1 on a set of realizations which cannot be covered by a category I set. This result assumes the continuum hypothesis.

Proposition 4 *There exists a distribution λ over the set of category tests T_C such that for every $\mu \in \Delta(\Delta(\Omega))$ with λ-probability one the non-expert fails the test with μ-probability one on a category II set of points in Ω. Hence*

$$\lambda(\{t \in T_C | \exists \text{ a category II set } S \subset \Omega \text{ such that } \mu(\{Q|S \cap t(Q) = \emptyset\}) = 1\}) = 1 \qquad (13)$$

Proof. A *Lusin set* L is an uncountable set such that every uncountable subset of L is of category II. The existence of such a subset of $[0, 1]$ was shown by Lusin (1914). In fact, every category II set contains a Lusin set (see Proposition 20.1 in Oxtoby (1980)). Furthermore, from Proposition 20.3 in Oxtoby (1980) we have that under the continuum hypothesis there exists a continuum of disjoint category II sets and each one of them is a Lusin set.

Denote the continuum of disjoint Lusin sets by F_α where α enumerates over the continuum. Fix an arbitrary test $t \in T_C$. For each α we define the test $t_\alpha \in T_C$ as follows:

$$t_\alpha(P) = (t(P) \setminus F_\alpha) \cup \{\omega \in F_\alpha | P(\{\omega\}) > 0\} \qquad (14)$$

We need to show that t_α as defined in (14) is indeed a category test. First note that if $t(P) \cap F_\alpha = \emptyset$ then $t_\alpha(P) = t(P)$. Otherwise, since F_α is a Lusin set we have that $t(P) \cap F_\alpha$ has at most a countable number of points since $t(P)$ is a category I set which implies that $t(P) \cap F_\alpha$ is category I. But as a category I subset of a Lusin set it must be countable. Hence

$$P(t(P)) = P(t(P) \setminus F_\alpha) + P(t(P) \cap F_\alpha) = P(t(P) \setminus F_\alpha) + \sum_{\{\omega \in F_\alpha | P(\{\omega\}) > 0\}} P(\{\omega\}) \qquad (15)$$

and we have that $P(t_\alpha(P)) = P(T(P)) = 1$ for all $P \in \Delta(\Omega)$ since the sum in (15) is countable. Note that $t(P) \setminus F_\alpha$ is always a measurable set since we remove a countable, hence measurable, set from the measurable set $t(P)$. Since $t_\alpha(P)$ is always a category I set (a union of a subset of a category I set with a countable set) we have that t_α is a member of T_C as required.

We first assume that for every α the set $\{\beta | t_\alpha = t_\beta\}$ is countable. Consider the set of tests $\{t_\alpha\}_\alpha$ as a measurable space induced by the mapping $f : [0, 1] \longrightarrow \{t_\alpha\}_\alpha$ where the continuum of indices α is taken from the unit interval $[0, 1]$ endowed with the Borel σ-field. As in the proof of Proposition 2 every non-atomic probability measure on $[0, 1]$ induces a non-atomic probability measure on $\{t_\alpha\}_\alpha$. Let λ be any non-atomic probability measure over the continuum of tests $\{t_\alpha\}_\alpha$ as derived from the unit interval (as before, every singleton is measurable in $\{t_\alpha\}_\alpha$). For any given measure $\mu \in \Delta(\Delta(\Omega))$ we have that with μ positive probability a test t_α is passed on at least one of the points in the set F_α only if $\bar{\mu}$ has at least one atom in F_α. This follows from noting that the test t_α at P includes a point from F_α if and only if that point is an atom of P. In other words, for μ to pass the test t_α we must have that for at least one $\omega \in F_\alpha$, $\bar{\mu}(\{\omega\}) > 0$ and so μ must assign positive probability to measures P with an atom at ω.

Since $\bar{\mu}$ has at most a countable number of atoms and the sets F_α are disjoint, we have that with positive μ-probability only a countable number of tests t_α can be passed, since the sets F_α are disjoint and passing each at one point from the set requires an atom in the set. Hence, with λ-probability 1 the randomized prediction μ will fail on a category II set of points.

The second case we need to consider is when there is a test $\bar{t}$ with an uncountable set of indices in $[0, 1]$ assigned to $\bar{t}$. The deterministic category test $\bar{t}$ cannot be manipulated on all but a category I set of realizations since a randomized prediction $\mu \in \Delta(\Delta(\Omega))$ would be required to have distinct atoms for each of the uncountable set of indices corresponding to $\bar{t}$.

The proof so far considered the set of realizations to be $[0, 1]$. To complete the proof we need to show that there exists a Lusin set in $\Omega = 2^{\aleph_0}$. This result is proven in the appendix. ■

The random category test that cannot be manipulated was constructed by randomly choosing among a continuum of tests. These tests were associated with large (category II) disjoint sets that can only be passed if the prediction has an atom in these sets (more precisely, the center of gravity of the prediction must have an atom there). Since at most a countable number of these conditions can be simultaneously passed by an uninformed expert, we have that any non-atomic measure over such a set of tests guarantees the inability to manipulate, and hence leads to failure on a category II set. We do not know if there is a deterministic category test that guarantees failure, even at a single realization.

4.1 Finitely determined tests

There is naturally an interest in finitely determined tests[5]. Finitely determined category tests t are those where $t(P)$ is a closed and nowhere dense set and hence for every realization ω not in $t(P)$ the test will fail in finite time. Obviously, these tests can only satisfy $P(t(P)) > 1-\varepsilon$. We conclude by showing that there is a randomization over finitely determined category tests that cannot be manipulated. Moreover, the randomization is over tests that share the same level of accuracy ε in determining the expert.

For every $t \in T_C$ and every $\varepsilon > 0$ we defined in Dekel and Feinberg (2004) a test $t_\varepsilon : \Delta(\Omega) \longrightarrow 2^\Omega$ such that $P(t_\varepsilon(P)) \geq 1-\varepsilon$ and $t_\varepsilon(P)$ is a closed and nowhere dense set for every $P \in \Delta(\Omega)$. Furthermore, $t_\varepsilon(P) \subset t(P)$ for all P, so for every ε and $t \in T_C$ there is a finitely determined test t_ε. These finitely determined tests are ε approximations of category tests.

Proposition 5 *There exists a distribution λ over T_C such that for every $\varepsilon > 0$ for every $\mu \in \Delta(\Delta(\Omega))$ with λ-probability 1 the prediction will fail the induced finitely determined randomly chosen test t_ε on an uncountable (or category II under the continuum hypothesis) set of points with μ-probability 1. Formally:*

$$\lambda(\{t \in T_C | \exists \text{ uncountable/category II set } S_t \subset \Omega \text{ such that } \mu(\{Q | S \cap t_\varepsilon(Q) = \emptyset\}) = 1\}) = 1 \tag{16}$$

Proof. Let λ be as in Proposition 2 (resp. Proposition 4 if the continuum assumption is made), that is, a probability measure over a class of category tests from T_C as a measurable space (where the class and measurable space are according to Propositions 2 and 4 respectively). Since $t_\varepsilon(P) \subset t(P)$ we have that

$$\{P \in \Delta(\Omega) | S_t \cap t_\varepsilon(P) = \emptyset\} \supset \{P \in \Delta(\Omega) | S_t \cap t(P) = \emptyset\}. \tag{17}$$

Applying (17) for the corresponding sets S_t—either $S_t = F_r$ when $t = t_r$ as in Proposition 2 or $S_t = F_\alpha$ when $t = t_\alpha$ as in Proposition 4—we have

$$\mu(\{P \in \Delta(\Omega) | S_t \cap t_\varepsilon(P) = \emptyset\}) \geq \tag{18}$$

$$\mu(\{P \in \Delta(\Omega) | S_t \cap t(P) = \emptyset\}) = 1 \tag{19}$$

[5] For example, calibration tests are defined as limits of such finitely observed events.

where the final equality follows from the corresponding proposition and choice of sets S_t. Hence

$$\lambda(\{t \in T_C | \mu(\{P \in \Delta(\Omega) | S_t \cap t_\varepsilon(P) = \emptyset\}) = 1\}) \geq \tag{20}$$

$$\lambda(\{t \in T_C | \mu(\{P \in \Delta(\Omega) | S_t \cap t(P) = \emptyset\}) = 1\}) = 1 \tag{21}$$

where the final equality follows from the corresponding proposition for the infinite tests above. ■

This uniform projection of a randomized category test to a randomization over finitely determined tests with the same level of accuracy demonstrates that no-manipulation results for infinite tests can be translated to no-manipulation results for finite approximations. Here, the finite approximation of a randomly chosen category test provides a test that identifies the non-expert in finite time and assures failure on a large set—an uncountable set, or even a category II set under the continuum hypothesis.

5 Appendix

Proof that there exists a Lusin set in Ω. The proof follows from viewing points in $\Omega = 2^{\aleph_0}$ as the dyadic (binary) expansion of points in $[0, 1]$. The dyadic expansion of the points in a Lusin set $L \subset [0, 1]$ must be a Lusin set in $\Omega = 2^{\aleph_0}$.

The dyadic expansion is unique for all but a countable set of points in $[0, 1]$. Assume by contradiction that the set of dyadic expansions of members of L, which we denote by $\bar{L}$, is not a Lusin set in Ω. Then we could find an uncountable category I subset of $\bar{L}$ in $2^{\aleph_0}$. It suffices to show that the inverse of the dyadic expansion maps a closed nowhere dense set in Ω to a closed nowhere dense set in $[0, 1]$ (hence a countable union of such sets will be mapped to at most a countable union of such sets). This will show that a category I set is mapped to a category I set and will contradict L being a Lusin set since the dyadic expansion and its inverse maps uncountable sets to uncountable sets.

Consider a closed set $S \subset \Omega$. Since S is closed under the product topology its map under the inverse of the dyadic expansion is closed; this is because convergence of the dyadic expansion implies convergence in $[0, 1]$. We need to show that if S is nowhere dense in Ω its preimage is nowhere dense in the interval. Consider any point in the interval and any open neighborhood of that point. Since the dyadic open intervals generate the same topology generated by open intervals we can find a dyadic interval in the open neighborhood which contains the point. The dyadic interval is open in Ω and hence contains points outside the nowhere dense set S. Hence these points are mapped in the inverse of the dyadic expansion to points in the dyadic interval. We conclude that every point in $[0, 1]$ has points from outside the image of S in any open neighborhood and the image of S is therefore nowhere dense as required. ■

References

[1] Dawid, A. P. (1982) "The Well-Calibrated Bayesian." *Journal of the American Statistical Association* **77** (379), 605–613.

[2] Dawid, A. P. (1985) "Calibration-Based Empirical Probability." *The Annals of Statistics* **13** (4), 1251–1274.

[3] Dekel, E., and Feinberg, Y. (2004) "Non-Bayesian Testing of an Expert." *Working paper.*

[4] Foster, D. P., and Vohra, R. V. (1998) "Asymptotic Calibration." *Biometrika* **85** (2), 379–390.

[5] Fudenberg, D., and Levine, D. K. (1999) "Conditional Universal Consistency." *Games and Economic Behavior* **29** (1-2), 104–130.

[6] Kalai, E., Lehrer, E., and Smorodinsky, R. (1999) "Calibrated Forecasting and Merging." *Games and Economic Behavior* **29** (1-2), 151–159.

[7] Lehrer, E. (2001) "Any Inspection Rule is Manipulable." *Econometrica* **69** (5) 1333–1347.

[8] Lusin, N. (1914) "Sur un problèmé de M. Baire ." *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Paris* **158** 1258-1261.

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

[10] Sandroni, A. (2003) "The Reproducible Properties of Correct Forecasts." *International Journal of Game Theory* **32** (1), 151–159.

[11] Sandroni, A., Smorodinsky, R., and Vohra, R. V. (2003) "Calibration with Many Checking Rules." *Mathematics of Operations Research* **28** (1), 141–153.