YOU WON'T HARM ME IF YOU FOOL ME

FEDERICO ECHENIQUE AND ERAN SHMAYA

Abstract. A decision maker faces a new theory of how certain events unfold over time. The theory matters for choices she needs to make, but possibly the theory is a fabrication. We show that there is a test which is guaranteed to pass a true theory, and which is also *conservative*: A false theory will only pass when adopting it over the decision maker's initial theory would not cause substantial harm; if the agent is fooled she will not be harmed.

We also study a society of conservative decision makers with different initial theories. We uncover pathological instances of our test: a society collectively rejects most theories, be they true or false. But we also find well-behaved instances of our test, collectively accepting true theories and rejecting false. Our tests build on tests studied in the literature on non-strategic inspectors.

Keywords: Conservatism, Testing, Strategic Expert, Merging of Opinions

1. Introduction

An agent makes decisions about events that unfold over time: to fix ideas, suppose she is trading in the stock market. The agent has a pre-existing belief about stock-market prices, beliefs which would lead her to adopt a certain contingent plan of actions (trades). Assume that the agent learns about a new theory of stock market prices. If she believes the theory, she should adopt a different contingent plan. The agent

Date: December, 2007.

Division of the Humanities and Social Sciences, California Institute of Technology and *MEDS, Kellogg School of Management, Northwestern University* Emails: fede@caltech.edu and e-shmaya@kellogg.northwestern.edu.

We thank Wojciech Olszewski for a comment on a previous draft. Echenique thanks the NSF for its support through award SES-0751980 and the Lee Center at Caltech.

faces a problem, should she adopt the theory or stick to her original beliefs?

We study agents who are *conservative*, in the sense that they are inclined to distrust the new theory, and follow their pre-existing beliefs. We show that, conservatism notwithstanding, agents can follow the new theory if they at the same time test it. When the theory is correct it is guaranteed to pass the test, and the agent reaps the benefits of following the optimal action. When the theory is incorrect, it will only pass the test when the resulting actions are not too suboptimal under the agent's pre-existing beliefs. We say "not too suboptimal" here to mean that in the limit, for arbitrarily patient agents, there is no loss in following an incorrect theory.

Thus the agent can be true to her conservative inclination, but follow the new theory. Under her pre-existing beliefs, she does not loose too much when the theory passes the test, even when the agent is *fooled*, i.e. when an incorrect theory passed the test. At the same time, she avoids the standard cost of conservatism—the cost of rejecting correct new theories. When the theory fails the test the agent knows that it is false. She can, for example, demand a restitution, or not trust the source of the theory in the future.

Our test has implications for a society of conservative agents with different initial beliefs. While each individual agent is assured that a true theory will pass her test with probability 1, it may be that for any given realization of stock market prices, a true theory would fail the tests of (in a sense) *most* agents. In fact, for some instances of conservative tests, most theories—whether true or false—fail most agents' tests at all outcomes.

We then present a conservative test that does not have this problem: A true theory will almost surely pass the tests of most agents. Moreover, there is an instance of conservative test which, in addition, is not manipulable: If the theory is a fabrication of a false expert then such a manipulation will not be successful.

Testing an expert. Our paper is closely related to the literature on testing experts. The expert presents a theory, and claims it is the true probability law governing some events that unfold over time; e.g. the weather or the stock market. The expert may be true, and report the actual law, or false, and be completely ignorant about the actual law. One would like to test the expert's report using the observed data. In certain environments, one cannot design a test which (a) the true expert is guaranteed to pass and (b) the false expert cannot manipulate (Foster and Vohra, 1998; Lehrer, 2001; Olszewski and Sandroni, 2007b, 2008; Shmaya, 2008).

We depart from the literature by modeling why one would want to test the expert: the expert's report matters for an agent's decisions. More specifically, we assume that the decision maker is engaged in a repeated situation of decision under uncertainty, so that in each period he has to choose an action and get payoffs that depend on the action and the realized state of nature. After seeing the report, the agent has to choose a contingent plan of actions. If she believes the expert, a certain contingent plan is optimal. If she does not believe him, an alternative plan, based on some given beliefs about the outcome, is optimal.

We show that there is a test such that (a) the true expert is guaranteed to pass and (b') the false expert can only pass when following his recommendation would not have lead to a significantly worse decision than ignoring it. So, while our test is manipulable in the sense that a false expert can easily pass it; the agent, if fooled into passing a false expert, is not harmed.

In addition, we consider an expert reporting to a society of agents (a guru). We show that there are tests satisfying (a) and (b') at the level of individual agents, but with very different implications for the society. One test fails most experts in the aggregate, whether they are true or not. Another test passes true experts in the aggregate, while being resistant to manipulation: A false guru can fool each agent separately, but he cannot fool a large set of agents.

The idea of using experts' prediction in decision making goes back at least to Hannan's no-regret theorem (Hannan, 1957). In Hannan's setup, a decision maker who is engaged in a repeated decision making situation gets advice from a finite set of experts. Hannan proves that the decision maker has a mixed strategy for choosing between the experts that guarantees that, in the long run, the DM will achieve no less than she could achieve by following a single expert in all periods.

Olszewski and Peski (2008) also consider an agent who uses an expert's advice in a decision-making model. In their model, the decision maker has no initial belief over the process of outcomes, but she has a default action she plans to take in the absence of any expert. They show that there is a contract between the decision maker and the expert that enables her to extract the full surplus of the expert's service: She always achieves at least as much as she would get if she followed her initial plan, and if the expert is true, she receives at least the payoff she would get if she knew the distribution and acted optimally.

We emphasize two major differences between our result and Hannan's no-regret theorem and Olszewksi and Peski's model. First, our model is more similar to the expert testing literature in that we reach a verdict about the expert's competence, passing a true expert almost surely. Second, our test is based only on the expert's prediction and the realized outcomes and is independent of the agent's payoff function. On the other hand, while Hannan's no-regret theorem and Olszewsik and Peski's model guarantee no regret over any realization of the underlying uncertainty, in our model the decision maker is not harmed according the her own initial beliefs.

Four additional papers are closely related to ours. First, Al-Najjar and Weinstein (2008) and Feinberg and Stewart (2008) study testing of multiple simultaneous experts. One can think of our agent's preexisting beliefs as a competing expert's theory. The closest paper is Al-Najjar and Weinstein's: they assume that one of the experts is true, and show that there is a test that only fails to single out the true expert when the different experts' predictions become close over time. This conclusion is similar to ours, in the sense that an agent following the recommendations of two experts who pass Al-Najjar and

Weinstein's test would eventually make similar decisions. In fact, we show in Section 4 that one instance of our test is formally similar to Al-Najjar and Weinstein's. One difference between the two is that, in our context, Al-Najjar and Weinstein's test is not guaranteed to pass a true expert. This issue is irrelevant in their model, as they want tests which single out a true expert when one is known to be present.

Finally, our tests in Section 4 build directly on the tests in Dekel and Feinberg (2006) and Olszewski and Sandroni (2007a); we show how one can combine the idea in our conservative tests with their tests to obtain tests with certain desirable properties in the society.

2. Model and Notation

The primitives of our model are (Z, A, r, λ, π):

- Z is a finite or countable set of *outcomes* equipped with the discrete topology; let $\Omega = Z^{\mathbb{N}}$ be the set of all sequences in Z.[1]
- A is a finite set of *actions*.
- $r : Z \times A \to [0, 1]$ is a *payoff function*.
- $\lambda \in (0, 1)$ is a *discount factor*.
- π is a probability measure over Ω; π represents given beliefs about Ω.

At every period n a decision maker (DM) chooses an action $a_n \in A$ and receives payoff $r(z_n, a_n)$, where z_n is the outcome of that period.

[1] We endow Ω with the product topology and the induced sigma algebra $\mathcal{B}$ of Borel sets.

The resulting discounted payoffs are

$$(1-\lambda)\sum_{n=0}^{\infty}\lambda^n r(z_n, a_n).$$

Let $Z^{<\mathbb{N}} = \bigcup_{n\in\mathbb{N}} Z^n$ be the set of finite sequences of elements of Z, including the empty sequence e. For $\omega = (z_0, z_1, \dots) \in \Omega$ and $n \in \mathbb{N}$, let $\omega|_n = (z_0, \dots, z_{n-1})$ be the initial segment of ω of length n. In particular $\omega|_0 = e$. For $s \in Z^{<\mathbb{N}}$ and $\omega \in \Omega$ we write $s \subseteq \omega$ if $s = \omega|_n$ for some n. For $s \in Z^{<\mathbb{N}}$ let $N_s = \{\omega \in \Omega | s \subseteq \omega\}$.

A (pure) strategy is given by $f : Z^{<\mathbb{N}} \to A$: $f(z_0, \dots, z_{n-1})$ is the action taken by the DM after observing $(z_0, \dots, z_{n-1})$.

For $\omega = (z_0, z_1, \dots) \in \Omega$ and a pure strategy $f : Z^{<\mathbb{N}} \to A$ let

$$R_\lambda(\omega, f) = (1-\lambda)\sum_{n\in\mathbb{N}}\lambda^n r\left(z_n, f(z_0, \dots, z_{n-1})\right) \tag{1}$$

be the discounted payoff to the DM who uses strategy f when the realization is ω. Say that f is ν-optimal iff

$$f \in \operatorname{argmax} \int R_\lambda(\omega, g)\nu(\mathrm{d}\omega),$$

where the maximization above is over strategies g.

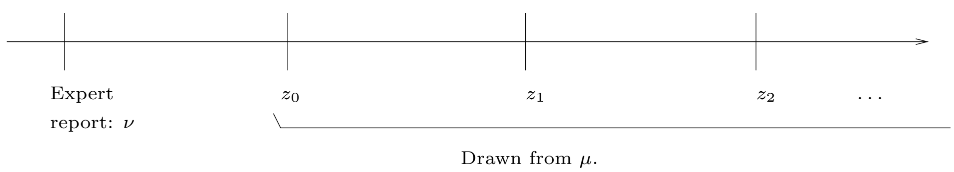
We denote the set of all probability measures over Ω by $\Delta(\Omega)$, and endow it with the weak* topology.

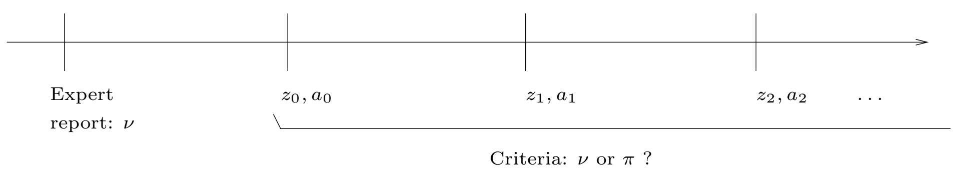
A probability measure $\nu \in \Delta(\Omega)$ is a *theory*. A *test function* is a function $T : \Delta(\Omega) \to 2^\Omega$: a theory ν is accepted if the observed realization of outcomes is in $T(\nu)$. A test function T is *type-I error free* if $\nu(T(\nu)) = 1$ for every $\nu \in \Delta(\Omega)$.

3. Individual Conservatism

We start with an interpretation of conservatism, then present our results.

Consider the first diagram in the figure below. Suppose the outcome of interest is a set Ω of infinite sequences of stock market prices, $z_0, z_1, \ldots$, drawn from a probability law μ. Before any z_n has been realized, an expert claims that the probability law ν governs the realization of prices over time; the expert's report may or may not coincide with μ. A test for the expert prediction is a set of outcomes $z_0, z_1, \ldots$ for which one decides that the expert has reported the true law.





We study the situation in the second diagram: A decision maker (DM) sees the expert prediction ν and has to choose, at each n an action a_n. DM has payoffs that depend on the sequences of outcomes, (z_n) and of actions (a_n). So she cares about the expert's report because it matters for the decision she has to make. If she believes the report is true, she should base her decisions on ν. If she believes it is false, she has some (given) pre-existing beliefs π about the outcomes.

DM is ignorant about μ so she has two criteria on which to base her choices: ν or π. A contingent plan f_ν is optimal if ν is true; f_π is optimal if DM rejects ν and sticks with π as the true theory. Evidently, f_ν is better than f_π if DM knew the expert's report to be true. Roughly speaking, our notion of conservatism is that the agent would instinctively prefer to stick with her initial belief and therefore follow f_π if the expert is false.

We show that there is a test such that, on outcomes for which ν passes the test, f_ν is at least as good as f_π under *either* of the two criteria DM might use.

One way to think of the result is that DM is conservative. She realizes that ν might be false, and is concerned about outcomes where f_ν leads to very different payoffs than f_π. She might want to reject ν on such outcomes. Our test assures DM that she can satisfy this conservative inclination, while getting the benefit of higher payoffs from following the true expert when ν is true.

Formally, we work with the following notion of conservatism.

Definition. Let $\pi \in \Delta(\Omega)$. A test function T is π-*conservative* if

$$\limsup_{\lambda \to 1} \int_{T(\nu)} \left(R_\lambda(\omega, g) - R_\lambda(\omega, f)\right)\ \pi(\mathrm{d}\omega) \leq 0 \tag{2}$$

for every $\nu \in \Delta(\Omega)$, every payoff function $r : Z \times A \to [0,1]$, and every ν-optimal strategy f and π-optimal strategy g, where R is given by (1).

Remark. Our definition of conservatism is independent of the payoff function r and of the optimal strategies f and g. For a fixed payoff function r, one might want to consider a test in which the decision maker accepts a theory on a sequence of outcomes by comparing the payoff she would get on that realization if she played according to some ν-optimal strategy f to the payoff she would get playing some π-optimal strategy g. Such a test might not satisfy (2) if we replace the optimal strategies f and g.

Theorem 1. *Let $\pi \in \Delta(\Omega)$, and let T_π be the test that is given by*

$$T_\pi(\nu) = \left\{ \omega \in \Omega \,\middle|\, \sup_n \frac{\pi(z_0) \cdot \pi(z_1|z_0) \cdot \ldots \pi(z_n|z_0, \ldots, z_{n-1})}{\nu(z_0) \cdot \nu(z_1|z_0) \cdot \ldots \nu(z_n|z_0, \ldots, z_{n-1})} < \infty \right\}, \tag{3}$$

where $\nu(z_n|z_0, \ldots, z_{n-1})$ and $\pi(z_n|z_0, \ldots, z_{n-1})$ are the forecasts made by ν and π about z_n given $z_0, \ldots, z_{n-1}$. Then T_π is π-conservative and type-I error free.

Remark. The test (3) is *prequential*, i.e. depends only on forecasts along $\omega = (z_0, z_1, \ldots)$.

Remark. The test (3) is similar to the test defined by Al-Najjar and Weinstein (2008), which accepts ν on

$$\left\{ \omega \in \Omega \,\middle|\, \sup_n \frac{\pi(z_0) \cdot \pi(z_1|z_0) \cdot \ldots \pi(z_n|z_0, \ldots, z_{n-1})}{\nu(z_0) \cdot \nu(z_1|z_0) \cdot \ldots \nu(z_n|z_0, \ldots, z_{n-1})} < 1 \right\}.$$

Note that their test can reject a true expert: If π is absolutely continuous w.r.t ν then the expert is rejected on the points where $1 < \frac{\mathrm{d}\pi}{\mathrm{d}\nu}$. On the other hand, Al-Najjar and Weinstein's test satisfies uniform convergence in ν in (2).

The proof of Theorem 1 is in Section 6, as are all proofs in the paper. The proof relies on the fact that for π-almost all outcomes on A, the posterior distributions of ν and π converge as n grows, and at the same time $\nu(T_\pi(\nu)) = 1$. We prove this assertion using Blackwell-Dubins merging theorem.

4. Implications of conservatism for the society.

In this section, we consider a society of decision makers, each has her own initial belief, and each uses a conservative test. We identify the decision maker with her initial belief, so the set of decision makers is given by $\Delta(\Omega)$.

We show that the collective response of the society to new theories can depend on the choice of individual tests. Concretely, we present three instances of conservative test. The first instance results in a society which accepts an expert only when the realization is an atom of his forecast: Note that at the individual level, a true theory is accepted with probability one. But in this instance of the test, the true expert is rejected, save for exceptional circumstances. Our second instance is a society which always accepts a true expert with probability 1, so the individual property of being type-I error free aggregates. Our third instance satisfies all the desirable properties of the second instance and, in addition, cannot be manipulated.

The tests in Instance 2 result from combining the ideas in our conservative tests with the test in Dekel and Feinberg (2006), while the tests in Instance 3 use the test in Olszewski and Sandroni (2007a).

We assume that every DM $\pi \in \Delta(\Omega)$ has a test $T_\pi : \Delta(\Omega) \to 2^\Omega$. To emphasize the dependence on π we sometimes use the term *individual test* for T_π. A collection $\{T_\pi\}_{\pi\in\Delta(\Omega)}$ of individual tests is called a *collective test*.

We use the notions of meager and residual sets as small and large sets, respectively: a set is *meager* if it is contained in a countable union of closed sets which has empty interior. A set is *residual* if its complement is meager.

Definition. Let $\nu \in \Delta(\Omega)$ be a theory and $\omega \in \Omega$ a realization. Then ν is *collectively accepted over* ω if the set of DM who accept ν, $P_{\nu,\omega} = \{\pi \in \Delta(\Omega)|\omega \in T_\pi(\nu)\}$, is residual. Say that ν is *collectively rejected over* ω if the set $P_{\nu,\omega}$ is meager.

Lemma 1. *Suppose that the individual tests T_π are type-I error free. Then for every theory ν and every atom ω of ν, ν is collectively accepted over ω.*

So an expert is always accepted by the society if the realization is an atom of his forecast. In particular, if ν is atomic then

$$\nu(\{\omega|\nu \text{ is collectively accepted over } \omega\}) = 1,$$

so a true expert that reports an atomic distribution is collectively accepted with probability 1.

Instance 1–The society rejects the true expert when he is not atomic. The following theorem shows the existence of society of conservative decision makers that collectively rejects all experts that do not gives a positive probability to the realized infinite sequence of outcomes.

Theorem 2. *Consider the collective tests T_π that are given in (3). Then a forecast $\nu \in \Delta(\Omega)$ is collectively accepted over a realization $\omega \in \Omega$ only if ω is an atom of ν.*

Remark. Recall that by Lemma 1, if the individual tests are type-I error free then a forecast $\nu \in \Delta(\Omega)$ is collectively accepted over a realization ω which is an atom of ν. Thus, among all the collective tests that are individually type-I error free, the test in Theorem 2 is the one that is the least favorable towards gurus, be they true or false.

Instance 2 – The society always accepts a true expert.

Theorem 3. *There exists a collective test $\{T_\pi\}_{\pi \in \Delta(\Omega)}$ such that*

(1) *For every $\pi \in \Delta(\Omega)$ the individual test T_π is conservative and type-I error free.*
(2) *A true expert is collectively accepted with probability 1:*

$$\nu\left(\{\omega | \nu \textit{ is collectively accepted over } \omega\}\right) = 1$$

for every $\nu \in \Delta(\Omega)$.

Instance 3 – Resistance to Strategic manipulation. Consider a false guru, who does not know the true distribution of the process. Such an expert can randomize a theory according to some probability distribution $\zeta \in \Delta(\Delta(\Omega))$. In the following theorem we describe a society that is immune to strategic manipulation by a false expert in the sense that over a topologically large set of realization such an expert will be almost surely rejected by the society.

Theorem 4. *There exists a collective test $\{T_\pi\}_{\pi\in\Delta(\Omega)}$ that satisfies the properties of Theorem 3 and, in addition, renders the society immune to strategic manipulation in the following sense: For every $\zeta \in \Delta(\Delta(\Omega))$, the set $\big\{\omega \big| \zeta(\{\nu|\nu \text{ is collectively rejected over } \omega\}) = 1\big\}$ is residual.*

Remark. Note that a specific decision maker with an initial belief π can be fooled by a false guru when he reports theory π (In which case she will be fooled but not harmed). However, the society described in Theorem 4 will collectively reject the false guru. He can fool one individual, but, generically, he cannot fool them all.

5. Concluding remarks

We present a collection of tests for conservative decision makers. The results reflect a basic criterion for conservatism: an inclination to distrust the proposed theory. We formalize this in a minimal and non-Bayesian model. The conservative criterion requires one to compare payoffs with and without following the proposed theory ν. The existence of an alternative π to ν is a requirement for modeling distrust of ν; what would one otherwise reject ν in favor of? Thus, it is natural for a conservative to evaluate ν according to π. We have discussed a class of tests (T_ν) which are satisfactory for conservatives because a false ν either fails T_ν or, *under* π, results in similar payoffs as π.

Note that our decision maker is not fully Bayesian: she does not have a prior belief about the correctness of the expert's theory. We simply explore the simple decisions between accepting and rejecting ν.

6. Proofs

6.1. **Preliminaries.**

6.1.1. *Lebesgue's decomposition.* Let π, ν be probability measures over Ω. π is *absolutely continuous w.r.t* ν ($\pi \ll \nu$) if $\nu(A) = 0$ implies $\pi(A) = 0$. π and μ are *singular* ($\pi \perp \nu$) if there exists $B \in \mathcal{B}$ such that $\nu(B) = 1$ and $\pi(B) = 0$.

Let $B \in \mathcal{B}$. Say that π is *absolutely continuous w.r.t.* ν *on* B (and write $\pi \ll_B \nu$) if $\nu(A) = 0$ implies $\pi(A \cap B) = 0$.

Proposition 5. *(Lebesgue Decomposition) Let ν and π be two probability measures. Then $\pi = \pi^{(r)} + \pi^{(s)}$ where $\pi^{(r)}$ and $\pi^{(s)}$ are finite measures such that $\pi^{(r)} \ll \nu$ and $\pi^{(s)} \perp \nu$.*

Corollary 6. *Let ν and π be probability measures, $\pi = \pi^{(r)} + \pi^{(s)}$ be the Lebesgue decomposition of π over ν, and let B be a set such that $\nu(B) = 1$ and $\pi^s(B) = 0$. Then $\pi \ll_B \nu$.*

Proof. For every $A \in \mathcal{B}$, $\nu(A) = 0$ implies

$$\pi(A \cap B) = \pi^{(r)}(A \cap B) + \pi^{(s)}(A \cap B) \leq \pi^{(r)}(A) + \pi^{(s)}(B) = 0,$$

as desired. $\square$

6.1.2. *Merging.* For a probability measure ν over Ω and $s \in Z^{<\mathbb{N}}$ we let ν_s be the probability measure over Ω that is given by

$$\nu_s(A) = \nu(A|N_s) = \nu(A \cap N_s)/\nu(N_s)$$

if $\nu(N_s) > 0$ and defined arbitrarily if $\nu(N_s) = 0$. The law of total expectation says that for every $n \in \mathbb{N}$ and every bounded Borel function R over Ω

$$\int R \,\mathrm{d}\nu = \int \left(\int R \,\mathrm{d}\nu_{\omega|_n} \right) \nu(\mathrm{d}\omega).$$

The distance between two probability measures ϕ_1 and ϕ_2 over Ω is given by

$$d(\phi_1, \phi_2) = \sup_{D \in \mathcal{B}} |\phi_1(D) - \phi_2(D)|.$$

Note that

$$\left| \int R \,\mathrm{d}\phi_1 - \int R \,\mathrm{d}\phi_2 \right| \leq d(\phi_1, \phi_2) \tag{4}$$

for every Borel function $R : \Omega \to [0, 1]$. Let π and ν be two probability measures over Ω. We say that ν *merges with* π if

$$\lim_{n \to \infty} d(\pi_{\omega|_n}, \nu_{\omega|_n}) = 0, \tag{5}$$

for π-almost every $\omega \in \Omega$. The following result was proved by Blackwell and Dubins.

Proposition 7. *Blackwell and Dubins (1962) If $\pi \ll \nu$ then ν merges with π*

Let $B \in \mathcal{B}$. Say that that ν *merges with* π *on* B if (5) is satisfied for π-almost every $\omega \in B$.

Corollary 8. *If* $\pi \ll_B \nu$ *then* ν *merges with* π *on* B.

Proof. If $\pi(B) = 0$ then the corollary is satisfied trivially. Assume $\pi(B) > 0$, and let π' be the probability measure over Ω that is given by $\pi'(D) = \pi(D|B) = \pi(D \cap B)/\pi(B)$ for every $D \in \mathcal{B}$. Then $\pi' \ll \nu$.

Let $s \in Z^{<\mathbb{N}}$ and $D \in \mathcal{B}$. Then it follows from the definition of π' that

$$\pi'(D|N_s) = \pi(D \cap N_s \cap B)/\pi(N_s \cap B) = \pi(D \cap B|N_s)/\pi(B|N_s).$$

It follows that

$$\pi'(D|N_s) \geq \pi(D \cap B|N_s) \geq \pi(D|N_s) - (1 - \pi(B|N_s)),$$

and

$$\pi'(D|N_s) \leq \pi(D|N_s)/\pi(B|N_s) \leq \pi(D|N_s) + (1 - \pi(B|N_s))/\pi(B|N_s)).$$

Therefore $d(\pi_s, \pi'_s) \leq (1 - \pi(B|N_s))/\pi(B|N_s)$. It follows that

$$\lim_{n\to\infty} d(\pi_{\omega|_n}, \pi'_{\omega|_n}) \leq \lim_{n\to\infty} (1 - \pi(B|N_{\omega|_n}))/\pi(B|N_{\omega|_n}) = 0 \tag{6}$$

π-almost surely on B since by the martingale convergence theorem $\pi(B|N_{\omega|_n}) \to \mathbf{1}_B(\omega)$.

By Blackwell-Dubins Theorem, since $\pi' \ll \nu$ it follows that

$$\lim_{n\to\infty} d(\pi'_{\omega|_n}, \nu_{\omega|_n}) = 0 \tag{7}$$

π'-almost surely on B. Since $\pi \ll_B \pi'$ it follows that (6) is satisfied π-almost surely on B. From (6) and (7) it follows that

$$d(\pi_{\omega|_n}, \nu_{\omega|_n}) \leq d(\pi_{\omega|_n}, \pi'_{\omega|_n}) + d(\pi'_{\omega|_n}, \nu_{\omega|_n}) \xrightarrow[n\to\infty]{} 0$$

π-almost surely on B, as desired. □

6.2. **Proof of Theorem 1.** We start by proving a proposition that describes a class of conservative tests.

Proposition 9. *Let $\pi \in \Delta(\Omega)$ and let $T : \Omega \to 2^\Omega$ be a test such that $\pi \ll_{T(\nu)} \nu$ for every $\nu \in \Delta(Omega)$. Then T is π-conservative.*

Remark. It follows from Proposition 9 and Corollary 6 that there exists a test which is π-conservative and type-I error free.

Proof of Proposition 9. By Bellman's principle of optimality, if f is ν-optimal then for every $s = (z_0, \ldots, z_{n-1})$ and every strategy g one has

$$\int_{N_s} R^n_\lambda(y, g)\nu_s(\mathrm{d}y) \leq \int_{N_s} R^n_\lambda(y, f)\nu_s(\mathrm{d}y)$$

where for every strategy h and every $\omega = (z_0, z_1, \ldots) \in \Omega$, $R^n_\lambda(y, h) = (1-\lambda)\sum_{k=n}^{\infty} r(z_k, h(z_0, \ldots, z_{k-1}))$ (an optimal strategy is optimal from every stage onward).

By Corollary 8, ν merges with π on $T(\nu)$. Let f be a ν-optimal strategy and g a π-optimal strategy. We claim that (2) is satisfied. Indeed, let $\varepsilon > 0$. Let $n \in \mathbb{N}$ be large enough such that $\pi\left(T(\nu) \setminus G\right) < \varepsilon$ where $G = T(\nu) \cap \left\{\omega \,\middle|\, d\left(\pi_{\omega|_n}, \nu_{\omega|_n}\right) < \varepsilon\right\}$, and let $0 < \Lambda < 1$ be large enough such that $(1 - \Lambda^n) < \varepsilon$. Let $\lambda > \Lambda$. For $\omega = (z_0, z_1, \ldots) \in \Omega$

and a strategy h let

$$R^n_\lambda(x,h) = (1-\lambda)\sum_{k=n}^{\infty} \lambda^k r\left(z_k, h(z_0,\ldots,z_{k-1})\right).$$

Then, by the choice of Λ and since $r(z,a)\in[0,1]$, it follows that

$$0 \le R^n_\lambda(\omega,h) \le R_\lambda(\omega,h) \le R^n_\lambda(\omega,h)+\varepsilon. \tag{8}$$

for every $\omega\in\Omega$ and every strategy h. Now,

$$\begin{aligned}
&\int_{T(\nu)} R_\lambda(\omega,g)\ \pi(\mathrm{d}\omega) < \int_G R_\lambda(\omega,g)\ \pi(\mathrm{d}\omega)+\varepsilon \le \int_G R^n_\lambda(\omega,g)\ \pi(\mathrm{d}\omega)+2\varepsilon = \\
&\int_G\left(\int R^n_\lambda(y,g)\ \pi_{\omega|_N}(\mathrm{d}y)\right)\pi(\mathrm{d}\omega)+2\varepsilon \le \int_G\left(\int R^n_\lambda(y,g)\ \nu_{\omega|_N}(\mathrm{d}y)\right)\pi(\mathrm{d}\omega)+3\varepsilon \le \\
&\int_G\left(\int R^n_\lambda(y,f)\ \nu_{\omega|_N}(\mathrm{d}y)\right)\pi(\mathrm{d}\omega)+3\varepsilon \le \int_G\left(\int R^n_\lambda(y,f)\ \pi_{\omega|_N}(\mathrm{d}y)\right)\pi(\mathrm{d}\omega)+4\varepsilon = \\
&\int_G R^n_\lambda(\omega,f)\ \pi(\mathrm{d}\omega)+4\varepsilon \le \int_{T(\nu)} R_\lambda(\omega,f)\ \pi(\mathrm{d}\omega)+4\varepsilon.
\end{aligned}$$

The first inequality follows from the choice of n and the fact that $R_\lambda(\omega,g)\le 1$. The second inequality follows from (8). The first equality follows from the law of total expectation. The third inequality follows from the definition of G and (4). The fourth inequality follows from the fact that f is ν-optimal and dynamic consistency. The fifth inequality follows from the definition of G and (4). The second equality follows from the law of total expectation. The last inequality follows from (8).

We proved that for every $\varepsilon>0$ there exists $0<\Lambda<1$ such that

$$\int_{T(\nu)} R_\lambda(\omega,g)\ \pi(\mathrm{d}\omega) < \int_{T(\nu)} R_\lambda(\omega,f)\ \pi(\mathrm{d}\omega)+4\varepsilon$$

for every $\lambda>\Lambda$. This completes the proof of the proposition. $\square$

□

Remark. Though our proof is based on Blackwell-Dubins merging theorem, a weaker notion of merging (Kalai and Lehrer, 1994) would have been sufficient for our cause.

Proof of Theorem 1. Fix $\pi, \nu \in \Delta(\Omega)$ and let $L^n : \Omega \to [0, \infty)$ be given by

$$L^n_{\pi,\nu}(\omega) = \frac{\pi(z_0) \cdot \pi(z_1|z_0) \cdot \ldots \pi(z_n|z_0, \ldots, z_{n-1})}{\nu(z_0) \cdot \nu(z_1|z_0) \cdot \ldots \nu(z_n|z_0, \ldots, z_{n-1})}.$$

So $L^n_{\pi,\nu}(\omega)$ is the likelihood ratio of π and ν over the realization $\omega|_n$.

By (Durrett, 1996, Chapter 4.3.c) $L^n_{\pi,\nu}$ is a martingale under ν, $\nu(L < \infty) = 1$, and the Lebesgue Decomposition (see Section 6.1.1) of π over ν is given by $\mathrm{d}\pi^{(r)} = L\mathrm{d}\nu$ and $\pi^{(s)}(A) = \pi(A \cap \{L = \infty\})$ where $L = \sup L^n_{\pi,\nu}$. In particular, it follows from Corollary 6 that $\pi \ll_{T(\nu)} \nu$. Therefore T_π satisfies the condition of Proposition 9, and so T_π is π-conservative and does not reject the truth with probability 1. □

6.3. **Proofs from Section 4.**

Proof of Lemma 1. Let $\nu \in \Delta(\Omega)$ be a forecast and let ω be an atom of ν. Since $\nu(T_\pi(\nu)) = 1$, it follows that $\omega \in T_\pi(\nu)$ for every π. Therefore $P_{\nu,\omega} = \Delta(\Omega)$. □

Proof of Theorem 2. Let $\nu \in \Delta(\Omega)$ be a theory and let $\omega \in \Omega$ be a realization such that $\nu(\{\omega\}) = 0$. We claim that ν is collectively rejected over ω. By (3)

$$P_{\nu,\omega} = \bigcup_{M=1}^{\infty} \bigcap_{n=1}^{\infty} \left\{\pi \in \Delta(\Omega) | \pi(N_{\omega|_n}) \leq M \cdot \nu(N_{\omega|_n})\right\}.$$

Since the sets $N_{\omega|n}$ are clopen it follows that the function $\pi \mapsto \pi(N_{\omega|n})$ is continuous in the weak* topology and therefore the sets $\left\{\pi \in \Delta(\Omega) | \pi(N_{\omega|n}) \leq M \cdot \nu(N_{\omega|n})\right\}$ are closed in the weak* topology and therefore $P_{\nu,\omega}$ is an F_σ set, i.e. a countable union of closed set.

We now claim that $P_{\nu,\omega}$ has an empty interior. Indeed, let $\pi \in P_{\nu,\omega}$, and let M be such that $\pi(N_{\omega|n}) \leq M \cdot \nu(N_{\omega|n})$ for every n. Then

$$\pi(\{\omega\}) = \lim_{n\to\infty} \pi(N_{\omega|n}) \leq M \cdot \lim_{n\to\infty} \nu(N_{\omega|n}) = M \cdot \nu(\{\omega\}) = 0.$$

Therefore $P_{\nu,\omega} \subseteq \{\pi \in \Delta(\Omega) | \pi(\{\omega\}) = 0\}$. The assertion that $P_{\nu,\omega}$ has empty interior follows from Lemma 2 below. □

Lemma 2. *For every $\omega \in \Omega$ the set $\{\pi \in \Delta(\Omega) | \pi(\{\omega\}) > 0\}$ is dense.*

Proof of Lemma 2. For every $\mu \in \Delta(\Omega)$ let $\pi_n = (1 - 1/n)\mu + 1/n\delta_\omega$ where δ_ω is dirac measure over ω. Then $\pi_n(\{\omega\}) > 0$ and $\pi_n \to \mu$ in the norm topology, and, in particular in the weak* topology. □

The proof of Theorem 3 uses the following proposition.

Proposition 10.
(1) *(Oxtoby, 1996, Theorem 16.5)For every $\nu \in \Delta(\Omega)$ there exists a meager subset F of Ω such that $\nu(F) = 1$.*
(2) *(Dekel and Feinberg, 2006, Proposition 1) For every meager subset F of Ω the set $\{\pi \in \Delta(\Omega) | \pi(F) > 0\}$ is a meager subset of $\Delta(\Omega)$.*

Proof of Theorem 3. For every $\nu \in \Delta(\Omega)$ let $t(\nu)$ be a meager subset of Ω such that $\nu(t(\nu)) = 1$. For every $\pi \in \Delta(\Omega)$ let $T_\pi : \Delta(\Omega) \to 2^\Omega$

be such that $T_\pi(\nu) = t(\nu)$ whenever $\pi(t(\nu)) = 0$ and $T_\pi(\nu)$ is defined arbitrary such that $\nu(T_\pi(\nu)) = 1$ and $\pi \ll_{T_\pi(\nu)} \nu$ whenever $\pi(t(\nu)) > 0$ (this is possible by Corollary 6). Then T_π does not reject the truth with probability 1, and, by Proposition 9 T_π is π-conservative.

Fix $\nu \in \Delta(\Omega)$ and let $\omega \in t(\nu)$. It follows from the definition of T that $\{\pi|\pi(t(\nu)) = 0\} \subseteq P_{\nu,\omega}$ and it follows from Proposition 10 that the set $\{\pi|\pi(t(\nu)) = 0\}$ is residual. Therefore the set $P_{\nu,\omega}$ is residual. It follows that ν is collectively accepted over ω for every $\omega \in t(\nu)$. Since $\nu(t(\nu)) = 1$ the result follows. □

The proof of Theorem 4 uses the following proposition proved by Olszewski and Sandroni (2007a).

Proposition 11. *There exists a function $t : \Delta(\Omega) \to 2^\Omega$ such that*

1. *For every $\nu \in \Delta(\Omega)$, $t(\nu)$ is meager and $\nu(t(\nu)) = 1$.*
2. *For every $\zeta \in \Delta(\Delta(\Omega))$ the set*

$$\left\{\omega\middle|\zeta(\{\nu|\omega \notin t(\nu)\}) = 1\right\}$$

is residual.

Proof of Theorem 4. Let $t : \Delta(\Omega) \to 2^\Omega$ be as in Proposition 11 and let $\{T_\pi\}_{\pi\in\Delta(\Omega)}$ be such that $T_\pi(\nu) = t(\nu)$ whenever $\pi(t(\nu)) = 0$ and $T_\pi(\nu)$ is defined arbitrary such that $\nu(T_\pi(\nu)) = 1$ and $\pi \ll_{T_\pi(\nu)} \nu$ whenever $\pi(t(\nu)) > 0$. Then T_π is a version of the test constructed in the proof of Theorem 3.

Fix a realization ω and let $\nu \in \Delta(\Omega)$ be such that $\omega \notin t(\nu)$. Then by the definition of T_π, $P_{\omega,\nu} \subseteq \{\pi|\pi(t(\nu)) > 0\}$. By Proposition 10 it follows that the set $\{\pi|\pi(t(\nu)) > 0\}$ is meager and therefore the set $P_{\omega,\nu}$ is meager. Thus, if $\omega \notin t(\nu)$ then $P_{\omega,\nu}$ is meager, i.e. ν collectively fails over ω. Let $\zeta \in \Delta(\Delta(\Omega))$. It follows from the previous observation that, for every realization ω,

$$\zeta(\{\nu|\omega \notin t(\nu)\}) = 1 \to \zeta(\{\nu|\nu \text{ is collectively rejected over } \omega\}) = 1.$$

Since the set $\left\{\omega\middle|\zeta(\{\nu|\omega \notin t(\nu)\}) = 1\right\}$ is residual by Proposition 11 it follows that the set $\left\{\omega\middle|\zeta(\{\nu|\nu \text{ is collectively rejected over } \omega\}) = 1\right\}$ is residual. $\square$

References

Al-Najjar, N., and J. Weinstein (2008): "Comparative Testing of Experts," *Econometrica*, 76(3), 541–560.

Blackwell, D., and L. Dubins (1962): "Merging of opinions with increasing information," *Annals of Mathematical Statistics*, 33, 882–886.

Dekel, E., and Y. Feinberg (2006): "Non-Bayesian testing of a stochastic prediction," *Review of Economic Studies*, 73.

Durrett, R. (1996): *Probability: Theory and Examples*. Wadsworth Publishing Company.

Feinberg, Y., and C. Stewart (2008): "Testing Multiple Forecasters," *Econometrica*, 76(3), 561–582.

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

Hannan, J. (1957): "Approximation to Bayes risk in repeated games," in *Contributions to the theory of games*, ed. by P. W. M. Dresher, A.W. Tucker. Princeton University Press.

Kalai, E., and E. Lehrer (1994): "Weak and Strong Merging of opinions," *Journal of Mathematical Economics*, 23, 73–86.

Lehrer, E. (2001): "Any Inspection Is Manipulable," *Econometrica*, 69(5), 1333–1347.

Olszewski, W., and A. Sandroni (2007a): "A NonManipulable Test," Forthcoming *Annals of Statistics*.

——— (2007b): "Strategic Manipulation of Empirical Tests," mimeo, Northwestern University.

——— (2008): "Manipulability of Future-Independent Tests," Forthcoming *Econometrica*.

Olszewski, W., and M. Peski (2008): "The principal-agent approach to testing experts," Mimeo Northwestern University.

Oxtoby, J. C. (1996): *Measure and Category: A Survey of the Analogies between Topological and Measure Spaces*. Springer.

Shmaya, E. (2008): "Many Inspections are Manipulable," *Theoretical Economics*, 3(3), 376–382.