

## Claim Validation<sup>†</sup>

By NABIL AL-NAJJAR, LUCIANO POMATTO, AND ALVARO SANDRONI\*

*Hume (1748) challenged the idea that a general claim (e.g., “all swans are white”) can be validated by empirical evidence, no matter how compelling. We examine this issue from the perspective of a tester who must accept or reject the forecasts of a potential expert. If experts can be skeptical about the validity of claims then they can evade rejection strategically. In contrast, if experts are required to conclude that claims backed by sufficient evidence are likely to be true, then they can be tested and rejected. These results provide an economic rationale for claim validation based on incentive problems. (JEL D82)*

A significant function of science, and of everyday thinking, is to make sense of the available evidence. This is often done by using facts to validate general claims (e.g., a pendulum 981 millimeters long *always* needs one second per swing). A difficulty, famously pointed out by David Hume (1748), is that examples of favorable evidence, however numerous, may not imply logically that a general claim must hold.

Hume’s work highlights different views on science. In one viewpoint, Nature works through laws and, ideally, science would uncover them with evidence. In another viewpoint, general claims are useful “as if models,” but without assurances that they are unbreakable, no matter how strong is the evidence in their favor.

Two hundred years after Hume’s work, Popper (1935) argued that what differentiates scientific claims from nonscientific claims is not validation by facts, but whether these claims can be falsified (i.e., rejected by the data). His leading example of a scientific statement was “all swans are white.” This example expresses the difference between validation and falsification. A Humean skeptic may doubt that the statement “all swans are white” is true, or likely to be true, even after many consecutive white swans are observed, but must agree that a single black swan

\*Al-Najjar: Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208 (e-mail: [al-najjar@northwestern.edu](mailto:al-najjar@northwestern.edu)); Pomatto: Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208 (e-mail: [l-pomatto@kellogg.northwestern.edu](mailto:l-pomatto@kellogg.northwestern.edu)); Sandroni: Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208 (e-mail: [sandroni@northwestern.edu](mailto:sandroni@northwestern.edu)). We thank three anonymous referees for many useful suggestions. We also thank Wojciech Olszewski, Eran Shmaya, Marciano Siniscalchi, and Rakesh Vohra for useful discussions, and seminar audiences at the Canadian Economic Theory Conference 2012, the Fifth Transatlantic Theory Workshop, the Summer meeting of the Econometric Society 2012, XIII Latin American Workshop in Economic Theory, Jolate conference in Bogota, and the Washington University seminar series. Sandroni gratefully acknowledges financial support from the National Science Foundation. All errors are ours.

<sup>†</sup> Go to <http://dx.doi.org/10.1257/aer.104.11.3725> to visit the article page for additional materials and author disclosure statement(s).

proves the statement false. Popper's argument is that if scientific statements are seen as testable hypothesis then they need not be validated. So, science can be confined to logic and evidence.

We examine the role of claim validation in an economic model where a self-proclaimed expert, named Bob, may misrepresent what he knows strategically. At period zero, Bob announces a finitely additive measure that can be used as a probabilistic guide to the future. As data unfold, Alice tests Bob's measure and decides whether to reject it. Alice faces an adverse selection problem. She does not know whether Bob knows something that she does not or if he is an uninformed expert who knows nothing relevant, but strategically may try to pass the test in order to maintain a false reputation of knowledge.

The measure announced by Bob may or may not be consistent with claim validation. This can be illustrated with the classic statement "all swans are white." A measure that satisfies claim validation is such that if many consecutive white swans are observed then "all swans are white" is deemed eventually a virtual certainty, provided that it is not *ex ante* ruled out as impossible.

The standard way to test a measure  $P$  is to define some event  $R^P$  that is unlikely to occur under the measure  $P$  and then reject it if  $R^P$  is observed. Any measure assigns low probability to some events. So, *any* measure, whether or not it validates claims, can be falsified. Thus, in the case of exogenously (or honestly) given measures, it is possible to speak, as Popper did, of falsification as a scientific method such that claim validation plays no significant role in it. As mentioned above, Popper portrayed falsification as a way to demarcate science from nonscience precisely because it is (seemingly) independent from claim validation. Now consider the case where Bob is (potentially) strategic.

Assume that Bob can announce any measure, whether or not it validates claims, and that he is completely uninformed about Nature. Then Bob has a mixed strategy that ensures that he is likely to pass Alice's test, regardless of how the data unfold. So, the feasibility of rejection is virtually removed. Now assume that Bob must announce a measure that validates claims. Then, there is a test such that no matter which strategy Bob uses, mixed or not, he is rejected by some data. Thus, the feasibility of rejection is preserved.

Incentives change the analysis of scientific methods. Assume that science is understood, like Popper and others did, as a collection of falsifiable claims without taking into account strategic considerations on how these claims are produced. Then, probabilistic claims can be falsified even if they are never validated. Now assume that probabilistic claims are produced by potentially strategic agents. Then, claim validation plays a significant role in the falsification of expert's claims.

Claim validation is notoriously difficult to justify. Yet, claim validation is relevant for a critical function of science: the falsification of potentially strategic expert's claims. These results provide an economic argument for claim validation based on incentive problems.

This paper is organized as follows: A brief literature review is in the next subsection. Main concepts are defined in Section I. Section II shows our main results. Section III concludes. Proofs are in the Appendix.

*Related Literature.*—A growing literature studies whether strategic experts can avoid rejection (see, among several contributions, Al-Najjar and Weinstein 2008; Al-Najjar et al. 2010; Babaioff et al. 2011; Cesa-Bianchi and Lugosi 2006; Chassang 2013; Dekel and Feinberg 2006; Feinberg and Stewart 2008; Feinberg and Lambert forthcoming; Fortnow and Vohra 2009; Foster and Vohra 1998; Fudenberg and Levine 1999; Gradwohl and Salant 2011; Gradwohl and Shmaya forthcoming; Hu and Shmaya 2013; Lehrer 2001; Olszewski and Peski 2011; Olszewski and Sandroni 2008, 2009a,b; Sandroni 2003; Sandroni, Smorodinsky, and Vohra 2003; Shmaya 2008; Stewart 2011; and Vovk and Shafer 2005). For a review, see Foster and Vohra (2013) and Olszewski (2011).

Our contribution utilizes mathematical techniques of this literature, but presents novel aspects. One difference is that our subject of analysis is the relationship between claim validation and testability. This motivation is not shared by any other paper we know of. The closest contribution to ours is Olszewski and Sandroni (2011), but their paper is not about claim validation. It is centered at the relationship between strict falsification and general falsification. In a follow-up paper (Pomatto, Al-Najjar, and Sandroni 2014), we analyze the connection between merging and testing opinions in a setting that rules out deterministic laws such as “all swans are white.” Chambers, Echenique, and Shmaya (2014) study the relation between falsifiability and the axiomatization of economic theories.

Unlike most of the literature, we consider a framework that allows for finitely additive measures. This is essential to model the antithesis of claim validation, Humean skepticism. Finitely additive measures satisfy de Finetti’s *coherence* principle (de Finetti 1974). Savage’s (1954) classic work imposes finite additivity, with additional restrictions. The relationship between finite additivity and inductive inference was formalized in Kelly (1996). Our attention to incentives and strategic motives has no counterpart in Kelly’s work. Gilboa and Samuelson (2012) show the positive effect of subjectivity on inductive inference.

## I. Basic Concepts

In every period, either 0 or 1 is observed. Let  $\Omega = \{0, 1\}^\infty$  be the set of all possible *paths*. So, a path  $\omega$  is an infinite sequence of outcomes. A path also expresses a general claim. If 1 is a white swan and 0 a nonwhite swan then the path  $(1, 1, 1, \dots)$  is the statement that all swans are white. Given a path  $\omega$  and a period  $t$ , let  $\omega^t \subseteq \Omega$  be the cylinder of length  $t$  with base  $\omega$ : i.e.,  $\omega^t$  are the paths that coincide with  $\omega$  in the first  $t$  periods. A cylinder  $\omega^t$  is the *data* available at time  $t$  along the path  $\omega$ . Let  $\mathcal{A}$  be any algebra containing all cylinders and singletons  $\{\omega\}$ , for all  $\omega \in \Omega$ . Let  $\Delta(\Omega)$  be the set of all finitely additive probability measures (henceforth, *measures*) on  $(\Omega, \mathcal{A})$ .<sup>1</sup> At period 0, a self-proclaimed expert, named Bob, delivers a measure  $P \in \Delta(\Omega)$  to a tester named Alice.

By conditioning on the data available at period  $t$ , the measure  $P$  gives odds of future events. In addition, Bob may or may not express skepticism about deducing

<sup>1</sup>Given a measurable space  $(X, \mathcal{A})$ , where  $X$  is a set and  $\mathcal{A}$  an algebra of subsets, a finitely additive probability measure is a function  $P : \mathcal{A} \rightarrow [0, 1]$  such that  $P(X) = 1$  and for every pair of disjoint events  $E_1, E_2$  in  $\mathcal{A}$ ,  $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ .

general claims from finite evidence. We now show that whether or not Bob expresses skepticism depends on the measure  $P$  that he communicates to Alice.

DEFINITION 1: A measure  $P \in \Delta(\Omega)$  validates claims if for every  $\omega \in \Omega$ ,

$$(1) \quad \lim_{t \rightarrow \infty} P(\omega^t) > 0 \Rightarrow \lim_{t \rightarrow \infty} P(\{\omega\} | \omega^t) = 1.$$

We denote by  $\Delta_v(\Omega)$  the set of measures that validate claims. Consider the classic statement that “all swans are white,” captured by the path  $1_\infty = (1, 1, \dots)$ . Consider a measure  $P$  that puts strictly positive probability on this path: i.e.,  $P(\{1_\infty\}) > 0$ . Then, after sufficiently many white swans are observed, the statement “all swans are white” becomes a virtual certainty. Thus, multiple repetitions of white swans validate the conclusion that all swans are white. More generally, assume that the odds of  $t$  consecutive white swans do not become vanishingly small as  $t$  goes to infinity. Under claim validation, the probability that “all swans are white” converges to 1 as increasingly many white swans are observed.

Consider the case in which Bob asserts that 0 or 1 have equal odds in every period, conditional on any history (i.e.,  $P$  corresponds to repeated tosses of a fair coin). In this example, the data  $\omega^t$  produced by *any* deterministic claim  $\omega$  has vanishing probability as  $t$  goes to infinity. The premise of (1) does not hold and so,  $P$  validates claims. If a measure validates claims then the data produced by some deterministic claim (e.g., “all swans are white”) must have bounded away from zero probability. We now show that while  $\sigma$ -additivity is commonly perceived as a technical axiom, it has deeper meaning.

REMARK 1: Any  $\sigma$ -additive measure  $P$  validates claims.

The main objective of this paper is to understand the role of claim validation. Thus, by Remark 1, measures must not be assumed, a priori, to be  $\sigma$ -additive. In contrast, some finitely additive measures validate claims and some do not. Hence, finite additivity is the proper assumption for our purposes.

We now sketch an example of a measure that does not satisfy claim validation, elaborating on de Finetti (1930) and Kelly (1996). The technical details are in the Appendix. Let  $1_t$  be a path where  $t$  consecutive white swans are followed by a nonwhite swan. Let  $P$  be a measure that assigns probability  $\varepsilon \in [0, 1]$  to the path  $1_\infty$  in which “all swans are white,” probability  $1 - \varepsilon$  to the event  $\{1_t : t \geq 1\}$  where a nonwhite swan appears at an unspecified point, and zero probability to each path  $1_t$ .<sup>2</sup> The measure has the property that no matter how many white swans are observed, a nonwhite swan may occur at an unspecified point, with nonvanishing probability. So, the conditional probability that “all swans are white” does not approach 1.

<sup>2</sup>The measure  $P$  can be constructed using only elementary mathematics.

### A. Empirical Tests

Alice announces her test at period 0, before Bob announces his measure. Given a measure  $P$  and a path  $\omega$ , the test returns a *pass* when the measure is accepted and a *fail* when the measure is rejected.

DEFINITION 2: A test is a function  $T : \Omega \times \Delta(\Omega) \rightarrow \{\text{fail}, \text{pass}\}$ .

The set  $A^P = \{\omega \in \Omega : T(\omega, P) = \text{pass}\}$  consists of all the paths deemed consistent with the announced measure  $P$ , while the *rejection set*  $R^P = (A^P)^c$  are the paths that reject  $P$ . We consider only tests that are measurable: i.e., such that  $A^P \in \mathcal{A}$  for every  $P \in \Delta(\Omega)$ . In a *rejection test*, the rejection set  $R^P$  is a countable union of cylinders (for every measure  $P$ ). So, rejection tests reject in finite time.<sup>3</sup> Whenever appropriate, we require the test to be a rejection test. Fix  $\epsilon \in [0, 1)$  and let  $\Lambda$  be a subset of  $\Delta(\Omega)$ .

DEFINITION 3: A test  $\Lambda$ -controls for type-I error with probability  $1 - \epsilon$  if for any  $P \in \Lambda$ ,

$$(2) \quad P(A^P) \geq 1 - \epsilon.$$

A test controls for type-I error if Bob considers likely to pass the test by truthfully announcing a measure he believes in. Given any measure  $P$  and  $\epsilon \in (0, 1)$  there is a set of cylinders  $R^P$  that have probability smaller than  $\epsilon$ . Thus, Bob's measure  $P$  can be tested and rejected whether or not it validates claims. In the absence of incentives problems, probabilistic claims can be feasibly falsified by the data whether or not claims are validated. We now examine the role of claim validation when there are incentive problems.

### B. Strategic Forecasting

An uninformed expert may try to pass the test strategically. He may pick his measures at random, but among finitely many measures.<sup>4</sup> Let  $\Delta_o\Delta(\Omega)$  be the set of measures with finite support on  $\Delta(\Omega)$ . An element  $\zeta \in \Delta_o\Delta(\Omega)$  is an expert's *strategy*.

DEFINITION 4: A test  $T$  can be manipulated with probability  $p \in [0, 1]$  if there is a strategy  $\zeta$  such that for every  $\omega \in \Omega$ ,

$$\zeta(\{P \in \Delta(\Omega) : \omega \in A^P\}) \geq p.$$

If a test is manipulable with high probability then a strategic expert is likely to pass the test, regardless of how the data unfolds and how much data is available.

<sup>3</sup>If  $T$  is a rejection test and  $\omega \in R^P$  there is a time  $t$  such that  $\omega^t \subseteq R^P$ .

<sup>4</sup>All the results extend to the case where  $\zeta$  is a  $\sigma$ -additive measure on  $\Delta(\Omega)$ .

DEFINITION 5: A rejection test  $T$  is nonmanipulable if for every strategy  $\zeta$  there is a cylinder  $C_\zeta$  such that for every path  $\omega \in C_\zeta$ ,

$$(3) \quad \zeta(\{P \in \Delta(\Omega) : \omega \in A^P\}) = 0.$$

A strategic expert cannot pass a nonmanipulable test for some finite realization of the data. So, nonmanipulable tests can discredit uninformed experts.

## II. Main Results

THEOREM 1: Consider the case where Bob can announce any measure in  $\Delta(\Omega)$ . Let  $T$  be a test that  $\Delta(\Omega)$ -controls for Type-I errors with probability  $1 - \epsilon$ . The test  $T$  can be manipulated with probability  $1 - \epsilon - \delta$ , for every  $\delta \in (0, 1 - \epsilon]$ .

THEOREM 2: Fix  $\epsilon \in (0, 1]$ . Consider the case where Bob's measure is required to validate claims. There exists a rejection test  $T$  that  $\Delta_v(\Omega)$ -controls for Type-I error with probability  $1 - \epsilon$  and is nonmanipulable.

Theorems 1 and 2 relate validation and falsification. Theorem 1 shows that if Bob can announce any measure (whether or not it validates claims) then he can avoid rejection strategically. It is nearly impossible to discredit him with any test that controls for type-I error, no matter how the data unfolds. This holds even if Alice could observe the entire infinite sequence of observations. Theorem 2 shows that if Bob is required to announce a measure that validates claims then he may be discredited, by some finite realization of the data.

These results show the transformative effect of incentives in the understanding of scientific methods. In the absence of incentive problems, any measure can be tested and rejected, even if they do not validate claims. Thus, claim validation is not essential to test and discredit honestly produced probabilistic claims. In contrast, in the presence of incentive problems, claim validation *is* relevant to discredit experts' probabilistic claims.

### A. Intuition behind the Proofs

The proof of Theorem 1 combines Fan's (1953) minmax, Banach-Alaoglu, and Riesz representation theorems. A test induces a zero-sum game between nature and the expert, such that nature's pure strategy is a path  $\omega \in \Omega$  and the expert's pure strategy is a measure  $P \in \Delta(\Omega)$ . The expert's payoff is 1 if his measure is accepted and 0 otherwise. The set of (mixed) strategies are respectively  $\Delta(\Omega)$  and  $\Delta_o\Delta(\Omega)$ . For every mixed strategy of nature, there exists a mixed strategy for the expert (to announce nature's measure) that gives him an expected payoff of  $1 - \epsilon$ . So, if the conditions of Fan's (1953) minmax theorem are satisfied, the expert obtains an expected payoff arbitrarily close to  $1 - \epsilon$ , for all paths  $\omega \in \Omega$ .

Fan's minmax theorem assumes (lower-semi) continuity of nature's payoff function and compactness of nature's mixed strategy space. The Banach-Alaoglu theorem and Riesz representation theorem deliver a topology that satisfies both conditions on  $\Delta(\Omega)$ , but not necessarily on subsets of  $\Delta(\Omega)$ , such as  $\sigma$ -additive measures.

In the testing strategic experts literature, measures are typically required to be  $\sigma$ -additive and impossibility results restrict Alice's tests (see, for example, Olszewski and Sandroni 2008 and Shmaya 2008). In contrast, in Theorem 1, there are no restrictions on tests (apart from control of type-I error).

The proof of Theorem 2 is based on a key property of measures that validates claims: for any such measure  $P$  it is unlikely that  $t_p$  consecutive white swans, followed by a nonwhite swan, occur, for  $t_p$  large enough. So, the measure  $P$  is rejected if a nonwhite swan appears after at least  $t_p$  successive white swans (see Olszewski and Sandroni 2011 for a related point). This test is nonmanipulable because all measures in the support of  $\zeta$  are rejected if a black swan appears after a long enough sequence of white swans.

The conclusions of the main results hold under weaker conditions. In Theorem 1, the assumption of a single test  $T$  can be relaxed. If Alice chooses randomly over different tests that  $\Delta(\Omega)$ -control for type-I errors, and the odds of the randomization are known to Bob, the expert can avoid rejection (see Theorem 12 in the online Appendix). In Theorem 2, it is enough to impose claim validation with respect to one fixed path. Fix a particular path  $\tilde{\omega}$  and say that a measure  $P$  validates claim  $\tilde{\omega}$  if (1) holds for path  $\tilde{\omega}$ . The proof of Theorem 2 shows that for every  $\epsilon > 0$  there exists a test that controls for type-I error with probability  $1 - \epsilon$  for all measures that validate claim  $\tilde{\omega}$  and is nonmanipulable.

### III. Conclusion

Traditional concepts such as claim validation and falsification can be formally studied in the context of a prediction problem. In the absence of incentive considerations, claim validation and falsification are not strongly related because probabilistic claims can be tested and rejected whether or not they are validated by evidence. In contrast, if an expert may misrepresent what he knows, then claim validation is relevant potentially to discredit uninformed experts' probabilistic claims. These results deliver an economic justification for claim validation. In the presence of incentive problems, claim validation plays a role in a critical function of science: the falsification of experts' probabilistic claims.

#### APPENDIX

##### A. Example

We now construct an example of a measure that violates claim validation, formalizing the discussion in the main text. For each  $t$ , denote by  $1_t$  a path where  $t$  initial white swans are followed by a nonwhite swan. To simplify the notation, denote by  $\bar{1}$  the path  $(1, 1, 1, \dots)$  where all swans are white. While our results hold for any algebra  $\mathcal{A}$  that contains all cylinders and all singletons, in this example we consider the special case of the smallest algebra containing the set  $\{1_t : t \geq 1\}$ , all cylinders and all singletons. We refer to this algebra as  $\mathcal{A}^s$ .

REMARK 2: Fix  $\epsilon \in (0, 1)$ . There exists a measure  $P$  defined on  $\mathcal{A}^s$  such that:

- (i)  $P(\{1_t : t \geq 1\}) = 1 - \epsilon$ ;
- (ii)  $P(\{\bar{1}\}) = \epsilon$ ; and
- (iii)  $P(\{1_t\}) = 0$  for every  $t$ .

By condition (i), Bob finds it likely that a nonwhite swan will eventually appear. By condition (ii), Bob considers possible that all swans are white. It now follows from (i)–(iii) that no matter how many white swans are observed, the conditional probability that “all swans are white” does not approach 1. To see this, note that

$$P(\{\bar{1}\} | \bar{1}^t) = \frac{P(\{\bar{1}\})}{P(\bar{1}^t)} = \frac{P(\{\bar{1}\})}{P(\{\bar{1}\}) + P(\{1_\tau : \tau \geq t\})} = \frac{\epsilon}{\epsilon + (1 - \epsilon)} = \epsilon$$

for every  $t$ , where the first equality follows from Bayes’ rule and the third equality follows from  $P(\{1_\tau : \tau \geq 1\}) = P(\{1_\tau : \tau \geq t\})$  for each  $t$ . Therefore,  $P$  does not validate claims. The intuition behind Remark 2 is simple: Conditions (1) and (2) determine directly the probability of the set  $\{1_t : t \geq 0\}$  and of each cylinder. Conditions (2) and (3) pin down the probability of each path. By additivity, we can derive the probability of all events in  $\mathcal{A}^s$ .

Algebras similar to  $\mathcal{A}^s$  are quite natural and suffice for defining and interpreting the concept of claim validation. While the algebra  $\mathcal{A}^s$  is useful to construct simple examples, larger algebras may have some appeal because they require Bob to deliver predictions on more events.

B. Proofs

PROOF OF REMARK 1:

Let  $P$  be  $\sigma$ -additive. Fix a path  $\omega$  such that  $\lim_{t \rightarrow \infty} P(\omega^t) > 0$ . Because  $\cap_{t \geq 1} \omega^t = \{\omega\}$ ,  $\sigma$ -additivity implies  $P(\omega^t) \rightarrow P(\{\omega\})$  as  $t \rightarrow \infty$  (see Shiryaev 1996). Therefore  $P(\{\omega\}) > 0$  and  $P(\{\omega\} | \omega^t) = \frac{P(\{\omega\})}{P(\omega^t)} \rightarrow 1$  as  $t \rightarrow \infty$ . Hence,  $P$  validates claims.

Our results will apply a known isomorphism between the strategic manipulation of tests and a certain class of zero-sum games between Nature and the expert. Given a test  $T$ , let  $\{A^P : P \in \Delta(\Omega)\}$  be the corresponding collection of acceptance sets and define the payoff function  $V : \Delta(\Omega) \times \Delta_o \Delta(\Omega) \rightarrow [0, 1]$  as

$$V(P, \zeta) = \sum_{Q \in \Delta(\Omega)} \zeta(Q) P(A^Q)$$

for every  $(P, \zeta) \in \Delta(\Omega) \times \Delta_o \Delta(\Omega)$ .



**THEOREM 3 (Fan 1953):** *Let  $X$  and  $Y$  be convex subsets of two vector spaces. Let  $f : X \times Y \rightarrow \mathbb{R}$ . If  $X$  is compact Hausdorff and  $f$  is concave with respect to  $Y$  and convex and lower semi-continuous with respect to  $X$  then*

$$(A1) \quad \min_{x \in X} \sup_{y \in Y} f(x, y) = \sup_{y \in Y} \min_{x \in X} f(x, y).$$

(Equation (A1) can have the form  $\infty = \infty$ . We refer the reader to Fan 1953 for a more general statement.) In order to apply Fan’s theorem we need to define a topology on  $\Delta(\Omega)$ . We endow  $\Delta(\Omega)$  with the weak\* topology. It can be defined as the coarsest topology that makes the function  $P \mapsto P(E)$  continuous for every event  $E \in \mathcal{A}$ . Under this topology,  $\Delta(\Omega)$  is Hausdorff. It follows from the Riesz representation theorem and the Banach-Alaoglu theorem that  $\Delta(\Omega)$  is compact (see Theorems IV.5.1 and V.4.2 in Dunford and Schwartz 1964).

**PROOF OF THEOREM 1:**

Consider the function  $V : \Delta(\Omega) \times \Delta_o\Delta(\Omega) \rightarrow [0, 1]$  defined above. The set  $\Delta(\Omega)$  is convex and compact. For every strategy  $\zeta$ , the map  $P \mapsto V(P, \zeta)$  is a convex combination of continuous functions, therefore it is continuous. Finally, the function  $V$  is affine in each variable. All of the conditions of Fan’s Minmax theorem are verified, so we can conclude that

$$(A2) \quad \min_{P \in \Delta(\Omega)} \sup_{\zeta \in \Delta_o\Delta(\Omega)} V(P, \zeta) = \sup_{\zeta \in \Delta_o\Delta(\Omega)} \min_{P \in \Delta(\Omega)} V(P, \zeta).$$

By assumption, the test satisfies  $P(A^P) \geq 1 - \epsilon$  for every  $P$ . Thus

$$\begin{aligned} \min_{P \in \Delta(\Omega)} \sup_{\zeta \in \Delta_o\Delta(\Omega)} V(P, \zeta) &\geq \min_{P \in \Delta(\Omega)} V(P, \delta_P) \\ &= \min_{P \in \Delta(\Omega)} P(A^P) \\ &\geq 1 - \epsilon \end{aligned}$$

where  $\delta_P(\{P\}) = 1$ .

Fix  $\delta > 0$ . By (A2) there exists a strategy  $\zeta \in \Delta_o\Delta(\Omega)$  such that  $V(P, \zeta) > 1 - \epsilon - \delta$  for every  $P \in \Delta(\Omega)$ . In particular,

$$V(\delta_\omega, \zeta) = \zeta(\{Q : \omega \in A^Q\}) \geq 1 - \epsilon - \delta$$

for every  $\omega \in \Omega$ , so the test can be manipulated with probability  $1 - \epsilon - \delta$ . This concludes the proof.

**PROOF OF THEOREM 2:**

We first show that a measure  $P$  validates claims if and only if it satisfies  $P(\omega^t) \rightarrow P(\{\omega\})$  as  $t \rightarrow \infty$  for every path  $\omega$ . To this end, suppose  $P$  validates claims and fix a path  $\omega$ . Define  $\delta = \lim_{t \rightarrow \infty} P(\omega^t)$ . If  $\delta = 0$  then  $P(\omega^t) \downarrow 0$ . Because

$P(\omega^t) \geq P(\{\omega\})$  for every  $t$ , then  $P(\{\omega\}) = 0$  and  $P(\omega^t) \rightarrow P(\{\omega\})$ . Now consider the case where  $\delta > 0$ . Since  $P$  validates claims, we have  $P(\{\omega\}|\omega^t) \rightarrow 1$ . Hence

$$P(\{\omega\}|\omega^t) = \frac{P(\{\omega\})}{P(\omega^t)} \rightarrow 1$$

so  $P(\omega^t) \rightarrow P(\{\omega\})$ . Conversely, assume  $P$  satisfies  $\lim_{t \rightarrow \infty} P(\omega^t) = P(\{\omega\})$  for every path  $\omega$ . Then,  $\lim_{t \rightarrow \infty} P(\omega^t) > 0$  if and only if  $P(\{\omega\}) > 0$ . In this case, by Bayes' rule,  $P(\{\omega\}|\omega^t) \rightarrow 1$  as  $t \rightarrow \infty$ . Therefore,  $P$  validates claims.

Fix  $\epsilon \in (0, 1)$  and a path  $\tilde{\omega}$ . For every measure  $P \in \Delta_v(\Omega)$  we have  $P(\tilde{\omega}^t) - P(\{\tilde{\omega}\}) = P(\tilde{\omega}^t - \{\tilde{\omega}\}) \rightarrow 0$  as  $t \rightarrow \infty$  (notice that  $\tilde{\omega}^t - \tilde{\omega}$  belongs to  $\mathcal{A}$  for every  $t$ ). Thus, for every  $P \in \Delta_v(\Omega)$  there exists a time  $t_P$  such that  $P(\tilde{\omega}^{t_P} - \{\tilde{\omega}\}) < \epsilon$ . For every  $P \in \Delta_v(\Omega)$  define the set  $\tilde{R}^P = \tilde{\omega}^{t_P} - \{\tilde{\omega}\}$ . Consider the test  $T$  where for every  $P \in \Delta_v(\Omega)$  the set  $\tilde{R}^P$  is the rejection set of  $P$ . By construction, the test  $\Delta_v(\Omega)$ -controls for Type-I error with probability  $1 - \epsilon$ . We now show it is nonmanipulable. Let  $(P_1, \dots, P_n)$  be the support of a strategy  $\zeta$ . Choose a period  $t_\zeta$  such that  $t_\zeta \geq t_{P_i}$  for  $i = 1, \dots, n$ . Then

$$(\tilde{\omega}^{t_\zeta} - \{\tilde{\omega}\}) \subseteq \bigcap_{i=1}^n \tilde{R}^{P_i},$$

hence  $\zeta(\{P \in \Delta(\Omega) : \omega \in A^P\}) = 0$  for every  $\omega \in (\tilde{\omega}^{t_\zeta} - \{\tilde{\omega}\})$ . The set  $\tilde{\omega}^{t_\zeta} - \{\tilde{\omega}\}$  is a union of cylinders. The proof is concluded by selecting, for every strategy  $\zeta$ , a cylinder  $C_\zeta \subseteq \tilde{\omega}^{t_\zeta} - \{\tilde{\omega}\}$ .

Let  $\mathcal{F}$  be the collection of all finite unions of cylinders, together with the empty set. This is the algebra generated by all cylinders. Denote by  $\tilde{\mathcal{A}}$  the algebra generated by  $\mathcal{F}$  and the collection of all singletons.

LEMMA 1: *A set  $E$  belongs to  $\tilde{\mathcal{A}}$  if and only if there exists a set  $F \in \mathcal{F}$  such that  $E \Delta F$  is finite.*

PROOF:

Let  $\mathcal{E}$  be the collection of all sets  $E$  such that there exists an event  $F \in \mathcal{F}$  for which  $E \Delta F$  is finite. By definition,  $\mathcal{F} \subseteq \tilde{\mathcal{A}}$  and  $\tilde{\mathcal{A}}$  contains all singletons. Therefore  $\mathcal{E} \subseteq \tilde{\mathcal{A}}$ . We now prove that  $\mathcal{E}$  is an algebra. Clearly  $\Omega, \emptyset \in \mathcal{E}$ . If  $E \in \mathcal{E}$  and  $E \Delta F$  is finite, because  $E^c \Delta F^c = E \Delta F$  and  $F^c \in \mathcal{F}$ , then also  $E^c$  belongs to  $\mathcal{E}$ . Let  $E_1, E_2 \in \mathcal{E}$ , and fix  $F_1, F_2 \in \mathcal{F}$  such that  $E_1 \Delta F_1$  and  $E_2 \Delta F_2$  are finite. Let  $E = E_1 \cup E_2$  and  $F = F_1 \cup F_2$ . Then  $E \Delta F \subseteq (E_1 \Delta F_1) \cup (E_2 \Delta F_2)$ . Hence  $E \Delta F$  is finite. Thus,  $\mathcal{E}$  is an algebra. Because it also contains all cylinders and all singletons,  $\tilde{\mathcal{A}} \subseteq \mathcal{E}$ . Hence  $\tilde{\mathcal{A}} = \mathcal{E}$ .

Given an event  $E \in \tilde{\mathcal{A}}$ , the set  $F \in \mathcal{F}$  such that  $E \Delta F$  is finite is unique. To see this, suppose  $F_1, F_2 \in \mathcal{F}$  are such that  $|E \Delta F_1|$  and  $|E \Delta F_2|$  are finite. Because  $F_1 \Delta F_2 = (E \Delta F_1) \Delta (E \Delta F_2)$ ,  $F_1 \Delta F_2$  is finite. Because the two sets are unions of finite cylinders,  $F_1 = F_2$ . We denote by  $F_E$  the unique event in  $\mathcal{F}$  such that  $E \Delta F_E$  is finite.

## PROOF OF REMARK 2:

Define on the algebra  $\mathcal{F}$  a measure  $Q$  as  $Q(F) = 1$  if  $\bar{1} \in F$  and  $Q(F) = 0$  otherwise. Now, define on the larger algebra  $\tilde{\mathcal{A}}$  the measure  $R$  as  $R(E) = Q(F_E)$ . For all  $E_1, E_2 \in \tilde{\mathcal{A}}$ ,  $F_{E_1 \cup E_2} = F_{E_1} \cup F_{E_2}$  and if  $E_1 \cap E_2 = \emptyset$  then  $F_{E_1} \cap F_{E_2} = \emptyset$ .<sup>5</sup> From these properties it follows that  $R$  is a well-defined measure. In particular, it satisfies  $R(\{\omega\}) = 0$  for every  $\omega$  and  $R(\bar{1}^t) = 1$  for every  $t$ . Let  $S = \epsilon \delta_{\bar{1}} + (1 - \epsilon) R$ , where  $\delta_{\bar{1}}(\{\bar{1}\}) = 1$ .

Define  $H = \{1_t : t \geq 1\} \cup \{\bar{1}\}$ . The algebra  $\mathcal{A}^s$  is the algebra generated by the collection  $\tilde{\mathcal{A}} \cup \{H\}$ . It is a standard result that a set  $E$  belongs to  $\mathcal{A}^s$  if and only if it is of the form  $E = (E_1 \cap H) \cup (E_2 \cap H^c)$  for some  $E_1, E_2 \in \tilde{\mathcal{A}}$ . Define a function  $P$  on  $\mathcal{A}^s$  as  $P((E_1 \cap H) \cup (E_2 \cap H^c)) = S(E_1)$  for every  $E_1, E_2 \in \tilde{\mathcal{A}}$ . It follows from the additivity of  $S$  that  $P$  is a measure. By construction, it satisfies  $P(\{\bar{1}\}) = S(\{\bar{1}\}) = \epsilon$ ,  $P(\{1_t : t \geq 1\}) = P(H) - P(\{\bar{1}\}) = 1 - \epsilon$ , and  $P(\{1_t\}) = S(\{1_t\}) = 0$  for every  $t$ .

## REFERENCES

- Al-Najjar, Nabil I., Alvaro Sandroni, Rann Smorodinsky, and Jonathan Weinstein. 2010. "Testing Theories with Learnable and Predictive Representations." *Journal of Economic Theory* 145 (6): 2203–17.
- Al-Najjar, Nabil I., and Jonathan Weinstein. 2008. "Comparative Testing of Experts." *Econometrica* 76 (3): 541–59.
- Babaioff, Moshe, Liad Blumrosen, Nicolas S. Lambert, and Omer Reingold. 2011. "Only Valuable Experts Can Be Valued." *Proceedings of the 12th ACM Conference on Electronic Commerce*, 221–22. San Jose, CA: ACM.
- Cesa-Bianchi, Nicolò, and Gábor Lugosi. 2006. *Prediction, Learning and Games*. New York: Cambridge University Press.
- Chambers, Cristopher P., Federico Echenique and Eran Shamaya. 2014. "The Axiomatic Structure of Empirical Content." *American Economic Review* 104 (8): 2303–19.
- Chassang, Sylvain. 2013. "Calibrated Incentive Contracts." *Econometrica* 81 (5): 1935–71.
- De Finetti, Bruno. 1930. "Sui Passaggi al Limite nel Calcolo delle Probabilità." *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere* 63: 1–12.
- De Finetti, Bruno. 1974. *Theory of Probability*. New York: Wiley.
- Dekel, Eddie, and Yossi Feinberg. 2006. "Non-Bayesian Testing of a Stochastic Prediction." *Review of Economic Studies* 73 (4): 893–906.
- Dunford, Nelson, and Jacob T. Schwartz. 1964. *Linear Operators, Part I*. New York: Interscience.
- Fan, Ky. 1953. "Minimax Theorems." *Proceedings of the National Academy of Science* 39 (1): 42–47.
- Feinberg, Yossi, and Nicholas S. Lambert. Forthcoming. "Mostly Calibrated." *International Journal of Game Theory*.
- Feinberg, Yossi, and Colin Stewart. 2008. "Testing Multiple Forecasters." *Econometrica* 76 (3): 561–82.
- Fortnow, Lance, and Rakesh V. Vohra. 2009. "The Complexity of Forecast Testing." *Econometrica* 77 (1): 93–105.
- Foster, Dean P., and Rakesh V. Vohra. 1998. "Asymptotic Calibration." *Biometrika* 85 (2): 379–90.
- Foster, Dean P., and Rakesh Vohra. 2013. "Calibration: Respite, Adspice, Prospice." In *Advances in Economics and Econometrics: Tenth World Congress*. Vol. 1, edited by Daron Acemoglu, Manuel Arellano, and Eddie, Dekel, 423–42. New York: Cambridge University Press.
- Fudenberg, Drew, and David K. Levine. 1999. "An Easier Way to Calibrate." *Games and Economic Behavior* 29 (1–2): 131–37.
- Gilboa, Itzhak, and Larry Samuelson. 2012. "Subjectivity in Inductive Inference." *Theoretical Economics* 7 (2): 183–215.
- Gradwohl, Ronen, and Yuval Salant. 2011. "How to Buy Advice." Unpublished.

<sup>5</sup>Let  $E = E_1 \cup E_2$  and  $F = F_{E_1} \cup F_{E_2}$ . Then  $E \Delta F \subseteq (E_1 \Delta F_{E_1}) \cup (E_2 \Delta F_{E_2})$  is finite. So  $F = F_E$ . Moreover,  $F_{E_1} \cap F_{E_2} \subseteq (E_1 \cup (F_{E_1} \setminus E_1)) \cap (E_2 \cup (F_{E_2} \setminus E_2))$ . The sets  $F_{E_1} \setminus E_1$  and  $F_{E_2} \setminus E_2$  are finite. So, if  $E_1 \cap E_2 = \emptyset$ , then  $F_{E_1} \cap F_{E_2}$  is finite, hence is the empty set.

- Gradwohl, Ronen, and Eran Shmaya.** Forthcoming. "Tractable Falsifiability." *Economics and Philosophy*.
- Hu, Tai Wei, and Eran Shmaya.** 2013. "Expressible Inspections." *Theoretical Economics* 8 (2): 263–80.
- Hume, David.** 1748. *An Enquiry Concerning Human Understanding*. London: Millar.
- Kelly, Kevin T.** 1996. *The Logic of Reliable Inquiry (Logic and Computation in Philosophy)*. New York: Oxford University Press.
- Lehrer, Ehud.** 2001. "Any Inspection Is Manipulable." *Econometrica* 69 (5): 1333–47.
- Olszewski, Wojciech.** 2011. "Calibration and Expert Testing." Unpublished.
- Olszewski, Wojciech, and Marcin Peski.** 2011. "The Principal-Agent Approach to Testing Experts." *American Economic Journal: Microeconomics* 3 (2): 89–113.
- Olszewski, Wojciech, and Alvaro Sandroni.** 2008. "Manipulability of Future-Independent Tests." *Econometrica* 76 (6): 1437–66.
- Olszewski, Wojciech, and Alvaro Sandroni.** 2009a. "A Nonmanipulable Test." *Annals of Statistics* 37 (2): 1013–39.
- Olszewski, Wojciech, and Alvaro Sandroni.** 2009b. "Strategic Manipulation of Empirical Tests." *Mathematics of Operations Research* 34 (1): 57–70.
- Olszewski, Wojciech, and Alvaro Sandroni.** 2011. "Falsifiability." *American Economic Review* 101 (2): 788–818.
- Pomatto, Luciano, Nabil Al-Najjar, and Alvaro Sandroni.** 2014. "Merging and Testing Opinions." *Annals of Statistics* 42 (3): 1003–28.
- Popper, Karl.** 1935. *The Logic of Scientific Discovery*. Republished 1968, London: Hutchinson & Co.
- Sandroni, Alvaro.** 2003. "The Reproducible Properties of Correct Forecasts." *International Journal of Game Theory* 32 (1): 151–59.
- Sandroni, Alvaro, Rann Smorodinsky, and Rakesh Vohra.** 2003. "Calibration with Many Checking Rules." *Mathematics of Operations Research* 28 (1): 141–53.
- Savage, Leonard, J.** 1954. *The Foundations of Statistics*. New York: Dover Publications.
- Shiryayev, Al'bert N.** 1996. *Probability*. New York: Springer-Verlag.
- Shmaya, Eran.** 2008. "Many Inspections Are Manipulable." *Theoretical Economics* 3 (3): 367–82.
- Stewart, Colin.** 2011. "Nonmanipulable Bayesian Testing." *Journal of Economic Theory*. 146 (5): 2029–41.
- Vovk, Vladimir G., and Glenn Shafer.** 2005. "Good Randomized Sequential Probability Forecasting is Always Possible." *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 67 (5): 747–63.