An Economic Model of Induction*

Nabil Al-Najjar † Luciano Pomatto ‡ Alvaro Sandroni §

April 23, 2013

Abstract

Inductive generalization (i.e., validation of general claims based on empirical evidence) is a critical method in science that is also notoriously difficult to justify. In particular, induction is not required to test honestly produced claims. We examine the role of induction in an economic model where agents may strategically misrepresent what they know. Our main result shows that induction *is* required to test and potentially reject expert's claims. This result provides an economic argument for induction based on incentive problems.

*We thank Wojciech Olszewski, Eran Shmaya, Marciano Siniscalchi and Rakesh Vohra for some useful discussions, as well as 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.

†Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208. (e-mail: al-najjar@kellogg.northwestern.edu)

‡Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208. (e-mail: l-pomatto@kellogg.northwestern.edu)

§Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208 (e-mail: sandroni@kellogg.northwestern.edu).

1 Introduction

A significant function of science, and of everyday thinking, is to make sense of available evidence. This is often done by validating general statements with facts. This critical process is known as inductive generalization and may be invoked when evidence backs a simple claim (e.g., a pendulum 981 millimeters long *always* needs one second per swing) or more complex ones such as the laws of supply and demand or those in classical mechanics. In this paper, we examine the role of induction in an economic model where statements may be strategically produced.

Induction has long been seen as an essential tool for the advancement of learning and also as a problematic method.[1] A practical difficulty with induction is that empirical regularities may be misleading (see Norton (2002)). For instance, Antoine Lavoisier found so many examples of oxygen in acids that he mistakenly considered it a universal law (and devised the name oxygen accordingly). Pierre de Fermat believed $2^{2^n} + 1$ to be a prime number. This claim holds for n smaller than 4294967297, but fails for that number.

While misleading empirical regularities show the practical risks of induction, the main difficulty with induction is about conceptually justifying the method itself. David Hume (1748) famously expressed skepticism about induction. His skepticism is summarized by Karl Popper (1979) who wrote:

> "Are we justified in reasoning from repeated instances of which we have experience to other instances of which we have no experience? Hume's unrelenting answer was: No, however great the number of repetitions."

The critical point in Hume's work is that an empirical justification for induction (e.g., induction has proven critical for the development of science) is a circular argument. So, in the absence of a logical justification for it, induction must either be accepted on faith or must be rejected. This is the problem of induction.

Two hundred years after Hume's work, Popper (1968, first published in 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 clearly the difference between induction 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 proves the statement false.

[1] Francis Bacon was one of the first to recognize the importance of induction. The key idea is that there are general statements such that all experimental results, under appropiate conditions, will always remain consistent with them. It simply cannot be different. Many scholars, such as Bernoulli, Keynes and Wittgenstein, have explored the concept of induction in different research fields.

Popper's essential claim is that as long as scientific statements are seen as testable hypothesis (never to be confirmed) then faith in induction is not necessary for science. Thus, science can be confined to logic and evidence. Popper's viewpoint that science is exclusively about falsification has been challenged, most notably by Thomas Kuhn (1962), but falsification remains a central method of science to this day.

The social implications of a deeper understanding of science can be substantial. When a group proclaims that their ideas are scientific and must be taught in public schools then the legitimacy of such demands must be examined under some conceptual evaluation on the nature of science. Analogously, the admissibility of testimony by expert witnesses must also be based on some evaluation of what is legitimate science.[2] A significant part of the economics profession embraced the positive view that science can be, at least conceptually, isolated from induction. For example, when an economist speaks of axiomatic decision theory having "empirical meaning," he/she does not imply that the axioms of decision theory have been confirmed by evidence, but rather that they can be refuted by data. However, whether it is really possible to conduce meaningful empirical analysis without appealing to induction has not been examined in an economic model. The purpose of this paper is to investigate whether empirical claims can be tested and rejected in the absence of faith in induction. That is, we ask whether induction must be presumed to hold in order to test experts' claims.

We approach this question in a formal economic model where a self-proclaimed expert, named Bob, may strategically misrepresent what he knows. At period zero, Bob announces a finitely additive measure that may be used as a guide to the future. As data unfold, Alice tests Bob's measure and decides whether to reject it or not. 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 may try to strategically pass the test in order to maintain a false reputation of knowledge.

In addition to providing forecasts about the future, Bob may or may not express faith in induction depending on the measure that he announces. This aspect of the model can be illustrated in the context of the classic statement "all swans are white." If Bob has faith in induction then sufficiently many consecutive white swans must convince him that "all swans are white," with high probability. So, we say that a measure is *inductive* if, conditional on sufficiently many white swans, the claim that all swans are white is a virtual certainty. If no matter how many white swans have been observed, Bob is still not virtually certain that "all swans are white" then he expresses Humean skepticism.

The standard way to test a measure P is to define some event R^P that is unlikely to

[2] See the ruling by U.S. District Court Judge William Overton on the teaching of intelligent design as science in *McLean* v. *Arkansas Board of Education*, Act 590 of the Acts of Arkansas of 1981 and the Daubert standard on science in *Daubert* v. *Merrell Dow Pharmaceuticals*, 509 U.S. 579.

occur under the measure P and then reject it if R^P is observed. Any measure P assigns low probability to some events R^P. So, *any* measure P, whether or not it expresses faith in induction, can be falsified. Thus, in the case of exogenously (or honestly) given measures, it is possible to speak, as Popper did, of falsification as an empirical method that does not necessitate induction. As mentioned above, Popper portrayed falsification as a promising way to demarcate science from non-science precisely because it is (seemingly) independent from induction. However, the connection between falsification and induction is much stronger when measures are produced by self-proclaimed experts who may strategically misrepresent their claims.

Assume that Bob is free to express Humean skepticism and that he is completely ignorant 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, Bob will not be discredited by any data. If skepticism about induction is allowed then strategic behavior virtually removes the feasibility of rejection. Now assume that Bob must announce a measure that expresses faith in induction. Then, there is a test such that no matter which strategy Bob uses (mixed or not) he is rejected by some realization of the data. Thus, Bob can be discredited by some data. If faith in induction is demanded then rejection is feasible.

These results highlight the transformative effects of incentives in the process of theory formation and in the analysis of scientific methods. Assume that science is understood, like Popper and many others did, as a collection of falsifiable claims without taking into account strategic considerations on how these claims are produced. In this popular view of science, induction and falsification are independent methods of empirical analysis because claims can be falsified in the absence (or presence) of faith in induction. Now assume that claims are produced by potentially strategic agents. Then, falsification requires an expressed faith in induction.

Popper proposed a solution to the problem of induction based on the argument that induction is not necessary for science because falsification demarcates science as realm different from nonscientific realms without logically pre-supposing induction. Kuhn provided a critique of Popper's criteria based on historical examples of scientific practices that are not entirely based on formulating and testing hypothesis. In contrast, our results show a logical justification for induction. In the presence of incentive problems, induction is essential for one of the critical functions of science: the testing and falsification of expert's claims.

This paper is organized as follows: A brief literature review is in section 1.1. Main concepts are defined in section 2. Section 3 presents our results in a simplified setting restricted to countably many paths. Section 4 shows the main results in a general setting. Section 5 concludes. Proofs are in the Appendix.

1.1 Related Literature

A growing literature studies whether strategic experts can avoid rejection (see, among several contributions, Al-Najjar and Weinstein (2008), Al-Najjar, Smorodinsky, Sandroni and Weinstein (2010), Babaioff, Blumrosen, Lambert and Reingold (2011), Cesa-Bianchi and Lugosi (2006), Chassang (2013), Dekel and Feinberg (2006), Feinberg and Stewart (2008), Feinberg and Lambert (2011), Fortnow and Vohra (2009), Foster and Vohra (1998), Fudenberg and Levine (1999), Gradwohl and Salant (2011), Gradwohl and Shmaya (2013), Lehrer (2001), Olszewski and Peski (2011), Olszewski and Sandroni (2008,2009a-b), Sandroni (2003), Shmaya (2008), Shmaya and Hu (2012), Stewart (2011), and Vovk and Shafer (2005)). For a review, see Foster and Vohra (2011) and Olszewski (2011).

Our contribution utilizes some the mathematical techniques of this literature, but also presents many novel aspects. The most salient difference is that our main subject of analysis is the relationship between induction and testability. This motivation is not shared by any other paper that we know of. The closest contribution to ours is Olszewski and Sandroni (2011), but their paper is not about induction. It is, instead, centered at the relationship between strict falsification (when an event deemed impossible occurs) and general falsification (based on the Cournot principle of rejection when an event deemed unlikely occurs). In a follow-up paper (Al-Najjar, Pomatto and Sandroni (2013)), we analyse the connection between merging and testing opinions in a setting that rules out deterministic laws such as "all swans are white." Partly related to this paper is also the work of Chambers, Echenique and Shmaya (2012), who study the relation between falsifiability and the axiomatization of economic theories.

Gilboa and Samuelson (2012) show a positive effect of subjectivity on inductive inference. Kelly (1996) relates induction with computability considerations and also formalizes the connection between induction and finitely additive measures. Our attention to incentives and strategic motives has no counterpart in Kelly's work.

2 Basic Concepts

2.1 Setup

In every period, either 0 or 1 is observed. The set $\{0,1\}^\infty$ consists of all possible *paths*. So, a path ω is an infinite history of outcomes. A path may also express a statement. As mentioned in the introduction, if 1 is a white swan and 0 a non-white swan then the path $(1,1,1,...)$ is the statement that all swans are white. Given a path ω and a period t, let $\omega^t \subseteq \{0,1\}^\infty$ be the finite cylinder of horizon t with base ω, i.e. ω^t is the set of paths that coincide with ω in the first t periods. A cylinder ω^t is the *data* available at time t along the path ω. Let Σ be a σ-algebra that contains all finite cylinders. Let Δ be the set of all finitely additive probability measures (henceforth, measures) on $(\{0,1\}^{\infty}, \Sigma)$.[3]

A measure P can be used as a guide to the future. Based on the available data, the conditional probabilities of P deliver odds of future events. At period 0, a self-proclaimed expert, named Bob, delivers a measure $P \in \Delta$ to a tester named Alice.

In the first part of the paper we assume, for expository simplicity, that Alice restricts her attention to a countable and dense subset Ω of $\{0,1\}^{\infty}$.[4] Accordingly, Bob must announce a measure that assigns probability 1 to Ω (i.e., a measure in $\Delta(\Omega)$). The general case is considered in section 4.

2.2 Induction

In addition to providing odds for the future, Bob may or may not express Humean skepticism about inductive generalizations (i.e., skepticism about deducing universal statements from finite data). We now show that whether or not Bob expresses Humean skepticism depends on the measure that he communicates to Alice.

Definition 1 A measure $P \in \Delta(\Omega)$ is *inductive* if for every path $\omega \in \Omega$,

$$\lim_{t \to \infty} P\left(\{\omega\} \mid \omega^t\right) = 1. \tag{1}$$

A measure is *skeptical i*f it is not inductive. Let $\Delta_i(\Omega)$ be the set of inductive measures.

Consider the classic statement that "all swans are white", captured by the path $1_{\infty} = (1, 1.....)$. If the measure P is inductive 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. This is the usual argument of enumerative induction.

It is straightforward to show that all $\sigma-$additive measures are inductive. So, while $\sigma-$additivity is commonly perceived as a technical axiom, it has deeper meaning. It presupposes induction. The main objective of this paper is to understand whether some scientific methods must pre-suppose induction. Thus, measures must not be assumed, a priory, to be $\sigma-$additive. In contrast, some finitely additive measures are inductive and some finitely

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

[4]Dense with respect to the product topology. That is, $\Omega \cap \omega^t \neq \emptyset$ for every finite cylinder ω^t. See Gilboa and Samuelson (2012) for a similar assumption.

additive measures are skeptical. Hence, finite additivity is the proper assumption for our purposes. Bob expresses *faith in induction* if he announces an inductive measure and *skepticism about induction* if he announces a skeptical measure.

We now sketch an example of a skeptical measure, elaborating on de Finetti (1930) and Kelly (1995). The details of this construction can be found in the Appendix. Let

$$1_t = \left(\underbrace{1, ..., 1}_{t}, 0, ...\right)$$

be a path in which the same outcome occurs for t periods and does not occur at period $t+1$. For example, this could be the path in which no human lives for 250 years until period t and at period $t+1$, a breakthrough occurs and someone does lives for 250 years. Let 1_∞ be the path in which "no human ever lives for 250 years." Let Q be the finitely additive measure that assigns (small) probability $\varepsilon > 0$ to the path 1_∞ while the remaining mass is "uniformly spread" over the set $\{1_t : t \geq 0\}$. While Q assigns zero probability to each event $\{1_t\}$, where the first person living 250 years occurs at precisely $t+1$, it also assigns probability $1-\varepsilon$ to event $\{1_t : t \geq 0\}$ where some people will live for 250 years at some unspecified point in the future. It can be easily shown that, no matter how many times people die younger than 250 years old, the conditional probability of the statement "no human will ever live for 250 years" does not approach 1. Hence, this finitely additive measure can express Humean skepticism that no matter how times the same outcome has been observed, an unprecedented event may still occur at some point in the future with non-vanishing probability.

2.3 Empirical tests

Alice announces her test at period 0, before Bob announces his measure. Given a measure P and a path ω, 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 $T : \Omega \times \Delta(\Omega) \longrightarrow \{\text{fail, pass}\}$.

The set $A^P = \{\omega \in \Omega : T(\omega, P) = \textit{pass}\}$ corresponds to all the paths deemed consistent with the announced measure P, while $R^P = \left(A^P\right)^c$ are the paths that reject P. In a *rejection test*, the rejection sets R^P is a union of finite cylinders (for every measure P). So, a rejection test rejects a measure in finite time.[5] Whenever appropriate, we require the test to be a rejection test. Fix $\epsilon \in [0,1)$ and let Λ be a subset of $\Delta(\Omega)$.

[5] So, if $\omega \in R^P$ then there is a time t such that $\omega^t \subseteq R^P$.

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

$$P\left(A^P\right) \geq 1-\epsilon. \tag{2}$$

A test controls for type-I error if whenever Bob believes that the odds of future events are accurately described by a measure P in Λ then he is also convinced that he is likely to pass the test by truthfully announcing what he believes.

Given any measure P and $\epsilon \in (0,1)$ there always exists a set of cylinders R^P that have probability smaller than ϵ. Thus, Bob's measure P can be tested whether it is inductive or skeptical. So, if Bob is assumed to honestly announce what he believes then there is no clear connection between testability and induction.

This basic notation suffices to express some of the central points mentioned in the introduction about induction and falsification in the absence of incentives problems. That is, induction and falsification are two prominent and quite distinct methodologies in science and in ordinary reasoning. Falsification seeks claims that can be discredited by the data. Induction seeks to validate general claims based on the available evidence. Exogenously given claims can be feasibly falsified by data whether or not induction is presumed to hold. We now examine the role of induction when claims are produced by potentially strategic experts.

2.4 Strategic forecasting

Condition (2) ensures that informed experts confidently expect a favorable verdict. In this section, we consider experts who are uninformed about the odds of future events, but may try to strategically pass the test. We allow strategic experts to use mixed strategies and pick their measures at random. However, we restrict randomization to be among finitely many measures. Let $\Delta_f\Delta(\Omega)$ be the set of measures with finite support on $\Delta(\Omega)$. An element $\zeta \in \Delta_f\Delta(\Omega)$ is called an expert's *strategy*.

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

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

If a test is manipulable with high probability then an uniformed, but strategic, expert is likely to pass the test regardless of how the data unfolds and how much data is available. Thus, if experts are potentially strategic then a meaningful empirical test must be non-manipulable.

Definition 5 A rejection test T is *non-manipulable* if for every strategy ζ there is a finite cylinder C_ζ such that for every path $\omega \in C_\zeta \cap \Omega$,

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

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

3 Induction and falsification

Proposition 1 *Consider the case where Bob can announce any measure (inductive or a skeptical) 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 $\delta \in (0, 1-\varepsilon]$.*

Proposition 2 *Fix $\epsilon \in (0,1]$. Consider the case where Bob is required to announce an inductive measure in $\Delta_i(\Omega)$. There exists a rejection test T that $\Delta_i(\Omega)-$controls for Type-I error with probability $1-\epsilon$ and is non-manipulable.*

Propositions 1 and 2 relate induction and falsification. Proposition 1 shows that if Bob is allowed to express skepticism about induction then he can strategically avoid rejection. So, it is near impossible to test and discredit him. Formally, *any* test that controls for type-I error with high probability can be passed with high probability, no matter how the data unfolds. This holds even if Alice could observe the entire infinite sequence of observations. Conversely, proposition 2 shows that if Bob is required to express faith in induction then he may be discredited, by some finite realization of the data. These results show that, under incentive problems, a more unified view of the basic methodologies of science (induction and falsification) emerges.

The proof of proposition 1 is a combination of the celebrated results of Fan's (1953) Minmax, Banach-Alaoglu, and Riez 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 $\Delta_f \Delta(\Omega)$. 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 expected payoff arbitrarily close to $1-\epsilon$, for all paths $\omega \in \Omega$.

A key condition in Fan's minmax theorem is the (lower-semi) continuity of Nature's payoff function and the compactness of Nature's mixed strategy space. The Banach-Alaoglu theorem and Riez representation theorem imply the existence of a topology on the (unrestricted) strategy space $\Delta(\Omega)$ such that both conditions are satisfied simultaneously. This is not necessarily true if attention is restricted to a subset of $\Delta(\Omega)$, such as those requiring induction.

The proof of proposition 2 is based on a crucial property of inductive measures. For every inductive measure P, there exists a large enough period t_P such that it is unlikely that t_P consecutive white swans, followed by a non-white swans, occur. So, an inductive measure P is rejected if a non-white swan appears after at least t_P successive white swans (see Olszewski and Sandroni (2011) for a related point). This test is non-manipulable because all measures in the support of ζ are rejected if a black swan appears after a large enough number of consecutive white swans.

4 Validation of candidate laws

We restricted attention to a countable set of paths. This is convenient for expositional reasons, but it is desirable to extend the analysis to a broader setting, where Bob can announce any measure $P \in \Delta$ on $(\{0,1\}^{\infty}, \Sigma)$. These are the probability assessments that satisfy de Finetti's *coherence* principle (de Finetti (1974)). So, we refer to this set-up as de Finetti's framework. (Savage (1954) classic work also imposes finite additivity, but makes additional restrictions on Δ).

It is straightforward to redefine the notions of tests and strategies in de Finetti's framework. A *test* is now a function $T : \{0,1\}^{\infty} \times \Delta \to \{\textit{fail}, \textit{pass}\}$ such that $T(\cdot, P)$ is Σ-measurable for every $P \in \Delta$. The definitions of strategies, Type-I errors and manipulability are obtained by replacing Ω with $\{0,1\}^{\infty}$ and $\Delta(\Omega)$ with Δ.

The simple induction of enumerative logic defined in (1) is not appropriate in de Finetti's framework. This follows because (1) requires positive mass in every path and this is only possible if the state-space is countable. A new definition of induction where $\lim_{t\to\infty} P(\{\omega\} \mid \omega^t) = 1$ whenever $P(\{\omega\}) > 0$ is appropriate if the set of states of the world is countable, but it is too weak in de Finetti's framework. Any non-atomic measure satisfies it automatically because, by definition, it assigns zero mass to all paths. Hence, inductive reasoning must go beyond simple enumeration. Instead, we consider validation of candidate laws. The intuition of this form of induction is as follows: suppose that after observing several consecutive white swans, all swans appearing in an increasingly large number of future periods are predicted to be white. If a measure is inductive in de Finetti's framework then, whenever the latter is true, the statement that "all swans are white" eventually becomes a virtual certainty. In order to formalize this concept, call a sequence of natural numbers $(\kappa_t)_{t=1}^{\infty}$ an *horizon* if it is increasing and unbounded.

Definition 6 A measure $P \in \Delta$ is *inductive in de Finetti's framework* if for every path $\omega \in \{0,1\}^{\infty}$, whenever

$$\lim_{t\to\infty} P\left(\omega^{t+\kappa_t}|\omega^t\right) = 1 \text{ for every horizon } (\kappa_t)_{t=1}^{\infty} \tag{4}$$

then

$$\lim_{t\to\infty} P\left(\{\omega\} \mid \omega^t\right) = 1.$$

Let Δ_i be the set of measures that are inductive in de Finetti's framework. Consider a candidate law ω. Assume that as data supporting this candidate law accumulates, Bob's predictions of finite future events becomes as if ω were a true law. That is, assume that condition (4) holds. Then, if Bob has faith in induction he is eventually almost certain that ω is a true law. So, laws are validated by supporting data, only when (4) holds. Analogous results to propositions 1 and 2 relating induction and testability hold in de Finetti's framework.

Theorem 1 Consider the case where Bob can announce any measure in Δ. Let T be a test that $\Delta-$controls for Type-I errors with probability $1-\epsilon$. The test T can be manipulated with probability $1-\epsilon-\delta$, for every $\delta \in (0, 1-\varepsilon]$. Thus, if Bob is not required to express faith in induction then he will not be discredited.

Theorem 2 Fix $\epsilon \in (0,1]$. Consider the case where Bob is required to announce an inductive measure in de Finetti's framework. There exists a rejection test T that Δ_i-controls for Type-I error with probability $1-\epsilon$ and is non-manipulable. Hence, in de Finetti's framework, induction is required for the feasibility of falsification.

Theorem 1 shows that if Bob is not required to express faith in induction then he will not be discredited. Theorem 2 shows that if Bob is required to express faith in induction then it is feasible to discredit him. Thus, induction is essential for falsification.

The intuition behind these results is similar to the intuition behind propositions 1 and 2. The first step in the proof of Theorem 2 is to show that a measure P is inductive if and only if it satisfies $\lim_{t\to\infty} P\left(\omega^t\right) = P\left(\{\omega\}\right)$ for every path ω. Once this critical property is demonstrated, the concluding argument follows the proof of proposition 2.

5 Conclusion

Traditional concepts such as induction and falsification can be formally studied in the context of a prediction problem. In the absence of incentive problems, induction and falsification are not strongly related because the claims of an expert can be tested and rejected whether or not induction is presumed to hold. However, if the expert may misrepresent what he knows then the expert must be required to express faith in induction in order to be potentially discredited. These results deliver a logical justification for induction. In the presence of incentive problems, induction is essential for a critical function of science: the falsification of expert's claims.

6 Appendix: Example

We now construct an example of a skeptical measure formalizing the discussion in the main text. Going back to definition (2), denote by 1_t the path where no one lives up to 250 years for the first t periods and then someone does live 250 years at period $t+1$. Let $1_\infty = (1, 1, 1, ...)$ be the path where no one ever lives for 250 years. Assume, without loss of generality, that all these paths belong to Ω. For every set $E \subseteq \{1_t : t \geq 0\}$, let l_E be the limit

$$\lim_{t \to \infty} \frac{|E \cap \{1_1, ..., 1_t\}|}{t}$$

whenever it exists. The limit l_E corresponds to the long run "frequency" of the event E. For instance, if $E = \{1_t : t \text{ is odd}\}$ then $l_E = \frac{1}{2}$. If E is finite, then $l_E = 0$ and if $E = \{1_t : t \text{ is a prime number}\}$ then $l_E = 0$. It can be shown that there exists a measure $U \in \Delta$ such that

$$U(E) = l_E$$

for every event E for which l_E is well-defined (see Rao and Rao (1982) for a proof). This measure is also known as a *Banach limit* in the repeated games literature. We can think about U as a uniform distribution over the set $\{1_t : t \geq 0\}$. Intuitively, under U, Bob is sure that the someone will live 250 years, but he is agnostic about when this remote event will occur. This is because for every fixed horizon $T < \infty$ we have $U(\{1_t : t < T\}) = 0$. In each of the first T period no one lives for 250 years almost surely.

To see how a skeptical measure can be obtained from U, consider now the mixture $Q = (1-\alpha) U + \alpha \delta_{1_\infty}$ where $0 < \alpha < 1$. For every $T < \infty$, denote by $\mathbf{1}_T = \{1_t : t > T\}$ the event corresponding to no one living 250 years in each of the first T periods. From the discussion above we conclude that

$$Q(\mathbf{1}_T) = (1-\alpha) U(\mathbf{1}_T) + \alpha \delta_{1_\infty}(\mathbf{1}_T) = 1$$

Therefore, even though $Q(\{1_\infty\}) > 0$, we have

$$Q(\{1_\infty\} | \mathbf{1}_T) = \frac{Q(\{1_\infty\})}{Q(\mathbf{1}_T)} = \frac{\alpha}{(1-\alpha)+\alpha} = \alpha$$

for every T, where the second equality follows from the fact that U is non atomic (hence $U(\{1_\infty\}) = 0$). Therefore Q is a skeptical measure.

7 Appendix: Proofs

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 $\{A^Q : P \in \Delta\}$ be the collection of its acceptance sets and define the payoff function $V : \Delta(\Omega) \times \Delta_f \Delta(\Omega) \longrightarrow \mathbb{R}$ as

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

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

Theorem 1 (Fan (1953)) *Let X and Y be convex subsets of two vector spaces.*[6] *Let $f : X \times Y \to \mathbb{R}$. If X is compact Hausdorff and f is concave with respect to Y and convex and lower semi-continuous with respect to X then*

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

In order to apply Fan's theorem we need to define a topology on Δ. We endow Δ 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. Under this topology, Δ is Hausdorff. It follows from the Riesz representation theorem and the Banach-Alaoglu theorem that Δ is compact (see IV.5.1 and V.4.2 in Dunford and Schwartz (1958)).

Proof of Proposition and Theorem 1. Consider the function $V : \Delta(\Omega) \times \Delta_f \Delta(\Omega) \longrightarrow \mathbb{R}$ defined above. The set $\Delta(\Omega) = \{P \in \Delta : P(\Omega) = 1\}$ is convex and closed, hence compact. For every strategy ζ, 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 the conditions of Fan's Minmax theorem are verified, so we can conclude that

$$\min_{P \in \Delta(\Omega)} \sup_{\zeta \in \Delta_f \Delta(\Omega)} V(P, \zeta) = \sup_{\zeta \in \Delta_f \Delta(\Omega)} \min_{P \in \Delta(\Omega)} V(P, \zeta) \tag{5}$$

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

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

[6] In Fan (1953), X and Y are not required to be subsets of vector spaces.

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

Fix $\delta > 0$. By (5) there exists a strategy $\zeta \in \Delta_f \Delta(\Omega)$ such that $V(P, \zeta) > 1 - \epsilon - \delta$ for every $P \in \Delta(\Omega)$. In particular,

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

for every $\omega \in \Omega$, so the test can be manipulated with probability $1 - \epsilon - \delta$. This concludes the proof of proposition 1. The proof of Theorem 1 follows the exact same argument. ■

Proof of Proposition and Theorem 2. We first prove Proposition 2. Fix a path $\omega \in \Omega$ and let $P \in \Delta_i(\Omega)$. Since $P(\{\omega\} | \omega^t) \to 1$ then $P(\omega^t) \to P(\{\omega\})$. Therefore, given $\epsilon > 0$, we can find a period t_P such that $P(\omega^{t_P} - \{\omega\}) < \epsilon$. Define the rejection set

$$R^P = \left(\omega^{t_P} - \{\omega\}\right) \cap \Omega.$$

By construction, the test induced by $\left\{R^P : P \in \Delta(\Omega)\right\}$ prevents Type-I errors for measures in $\Delta_i(\Omega)$ with probability $1 - \epsilon$.

Now, let $(P_1, ..., P_n)$ be the support of a strategy ζ. Choose a period t_ζ such that $t_\zeta \geq t_{P_i}$ for $i = 1, ..., n$. Then

$$\left(\omega^{t_\zeta} - \{\omega\}\right) \cap \Omega \subseteq \cap_{i=1}^n R^{P_i}$$

Since Ω is dense, the set $(\omega^{t_\zeta} - \{\omega\}) \cap \Omega$ is nonempty and $\zeta\left(\left\{P \in \Delta(\Omega) : \omega \in A^P\right\}\right) = 0$ for every $\omega \in (\omega^{t_\zeta} - \{\omega\}) \cap \Omega$. The proof of proposition 2 is concluded by selecting a cylinder C_ζ in $\omega^{t_\zeta} - \{\omega\}$ for every ζ.

In order to prove theorem 2, we first show that if a measure P belongs to Δ_i then $P(\omega^t) \to P(\{\omega\})$ for every $\omega \in \{0,1\}^\infty$. To this end, let $P \in \Delta_i$ and assume by way of contradiction that there exists a path $\omega \in \{0,1\}^\infty$ such that $P(\omega^t) \nrightarrow P(\omega)$. By defining $\gamma = \inf_t P(\omega^t) > 0$, we have $P(\omega^t) = \gamma + \xi_t$ for some sequence $(\xi_t)_{t=1}^\infty$ such that $\xi_t \downarrow 0$. For every t and $\kappa \in \mathbb{N}$,

$$P\left(\omega^{t+\kappa} | \omega^t\right) = \frac{P\left(\omega^{t+\kappa}\right)}{P\left(\omega^t\right)} = \frac{\gamma + \xi_{t+\kappa}}{\gamma + \xi_t} \geq \frac{\gamma}{\gamma + \xi_t}$$

In particular, for every horizon $(\kappa_t)_{t=1}^\infty$

$$P\left(\omega^{t+\kappa_t} | \omega^t\right) \geq \frac{\gamma}{\gamma + \xi_t}$$

therefore $P(\omega^{t+\kappa_t} | \omega^t) \to 1$ as $t \to \infty$. Since P belongs to Δ_i, it follows that $P(\{\omega\} | \omega^t) \to 1$, hence $P(\omega^t) \to P(\{\omega\})$. A contradiction. Therefore, we can conclude that $P(\omega^t) \to P(\omega)$ for every $\omega \in \{0,1\}^\infty$ and every measure P in Δ_i. We can now proceed along the lines of the proof of proposition 2. Fix a path $\widetilde{\omega}$ and construct a test where for every measure $P \in \Delta_i$ the rejection set is defined as $R^P = \widetilde{\omega}^{t_P} - \{\widetilde{\omega}\}$, for some t_P such that $P\left(\widetilde{\omega}^{t_P} - \{\widetilde{\omega}\}\right) < \epsilon$. The test is non manipulable. This concludes the proof of theorem 2. ■

References

[1] Al-Najjar, N., L. Pomatto, A. Sandroni (2013): "Merging and Testing Opinions," *mimeo.*

[2] Al-Najjar, N., A. Sandroni, R. Smorodinsky and J. Weinstein (2010): "Testing Theories with Learnable and Predictive Representations," *Journal of Economic Theory*, 145(6), 2203-2217.

[3] Al-Najjar, N., and J. Weinstein (2008): "Comparative Testing of Experts," *Econometrica*, 76(3), 541-559.

[4] Babaioff, M., L. Blumrosen, N. Lambert and O. Reingold (2011): "Only Valuable Experts Can Be Valued," *Proceedings of the 12th ACM conference on electronic commerce*, 221-222.

[5] Cesa-Bianchi, N. and G. Lugosi (2006): *Prediction, Learning and Games*, Cambridge University Press.

[6] Chambers, C., F. Echenique and E. Shamaya. (2012): "The Axiomatic Structure of Empirical Content," *mimeo.*

[7] Chassang, S. (2013): "Calibrated Incentive Contracts," *Econometrica*, forthcoming.

[8] de Finetti, B. (1930): "Sui Passaggi al Limite nel Calcolo delle Probabilita," *Rendiconti del Reale Istituto Lombardo di Scienze e Lettere*, 63, 1-12.

[9] de Finetti, B. (1974): *Theory of Probability*, Wiley, New York.

[10] Dekel, E. and Y. Feinberg (2006): "Non-Bayesian Testing of a Stochastic Prediction," *Review of Economic Studies*, **73**, 893 - 906.

[11] Dunford, N., and J. T. Schwartz (1964): *Linear Operators, Part I*, Interscience, New York.

[12] Fan, K. (1953): "Minimax Theorems," *Procedings of the National Academy of Science*, 39(1), 42-47.

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

[14] Feinberg, Y., and N. Lambert (2011): "Mostly Calibrated," *mimeo.*

[15] Fortnow, L., and R. Vohra (2009): "The Complexity of Forecast Testing," *Econometrica*, 77(1), 93-105.

[16] Foster, D. P., and R. Vohra (1998): "Asymptotic Calibration," *Biometrika*, 85(2), 379-390.

[17] Foster, D. P., and R. Vohra (2011): "Calibration: Respice, Adspice, Prospice," *mimeo.*

[18] Fudenberg, D., and D. Levine (1999): "An Easier Way to Calibrate," *Games and Economic Behavior*, 29(1-2), 131-137.

[19] Gilboa, I., and L. Samuelson (2012): "Subjectivity in Inductive Inference," *Theoretical Economics*, 7, 183-215.

[20] R. Gradwohl and Y. Salant (2011): "How to Buy Advice," *mimeo.*

[21] R. Gradwohl and E. Shmaya (2013): "Tractable Falsification," *mimeo.*

[22] Hu, T., and E. Shmaya (2012): "Expressible Inspections," *Theoretical Economics*, forthcoming.

[23] Hume, D. (1748): *An Enquiry Concerning Human Understanding.*

[24] Kelly, K. T. (1996): "The Logic of Reliable Inquiry," Oxford University Press, New York.

[25] Kuhn, T. (1962): *The Structure of Scientific Revolutions*, The University of Chicago Press.

[26] Lehrer, E. (2001): "Any Inspection is Manipulable," *Econometrica*, 69(5), 1333-1347.

[27] Norton, J. D. (2010): "A Survey of Inductive Generalization," *mimeo.*

[28] Olszewski, W. (2011): "Calibration and Expert Testing," *mimeo.*

[29] Olszewski, W., and A. Sandroni (2008): "Manipulability of Future Independent Tests," *Econometrica*, 76, 1437 - 1466.

[30] Olszewski, W., and A. Sandroni (2009a): "A Nonmanipulable Test," *Annals of Statistics*, 37, 1013-1039.

[31] Olszewski, W., and A. Sandroni (2009b): "Strategic Manipulation of Empirical Tests," *Mathematics of Operations Research*, 34, 57-70.

[32] Olszewski, W., and A. Sandroni (2011): "Falsifiability," *American Economic Review*, 101, 788-818.

[33] Olszewski, W. and M. Peski (2011): "The Principal-Agent Approach to Testing Experts," *American Economic Journal: Microeconomics* 3, 89-113

[34] Popper, K. (1968): "The Logic of Scientific Discovery," Hutchinson Co, London.

[35] Popper, K. (1979): "Objective Knowlege: an Evolutionary Approach," Oxford University Press, Oxford.

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

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

[38] Shmaya, E. (2008): "Many Inspections are Manipulable," *Theoretical Economics*, 3, 367-382.

[39] Stewart, C. (2011): "Non-manipulable Bayesian Testing," *Journal of Economic Theory*, 146(5), 2029-2041.

[40] Vovk, V., and G. Shaffer (2005): "Good Randomized Sequential Probability Forecasting is Always Possible," *Journal of the Royal Statistical Society Series B*, 67(5), 747-763.