Falsifiability*

Wojciech Olszewski[†] and Alvaro Sandroni[‡]

December 19, 2009

Abstract We examine Popper's falsifiability within an economic model in which a tester hires a potential expert to produce a theory. Payments are contingent on the performance of the theory vis-a-vis data. We show that if experts are strategic, falsifiability has no power to distinguish scientific theories from worthless theories. The failure of falsification in screening informed and uninformed experts motivates questions on the broader concepts of refutation and verification. We demonstrate an asymmetry between the two concepts. Like falsification, verification contracts have no power to distinguish between informed and uninformed experts, but some refutation contracts are capable of screening experts.

The publication of "*The Logic of Scientific Discovery*" by Karl R. Popper (1968, first published in 1935) was a transformative event in the philosophy of science because it expressed clearly the concept of falsifiability. Popper was interested in demarcation criteria that differentiate scientific ideas from nonscientific ideas (and hence give

*We are also grateful to Nabil Al-Najjar, Sandeep Baliga, Eddie Dekel, Yossi Feinberg, Roger Myerson, and Rakesh Vohra for comments and conversations. We thank workshop participants at Brown, Caltech, Columbia, Cornell, Yale, Harvard - MIT, Northwestern, Penn State, Princeton, Stanford - GSB, Rice, University of Bonn, University of British Columbia, University of California at Berkeley, UCLA, University of Chicago, Instituto de Matematica Pura e Aplicada (IMPA), University of Texas at Austin, University of Toronto, Washington University at St. Louis, and at the North American Summer Meeting of the Econometric Society in Minneapolis in 2006. We also thank the National Science Foundation (awards CAREER SES - 0644930; SES - 0820472 and SES-0922404, respectively) for financial support.

[†]Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston IL 60208; e-mail: wo@northwestern.edu

[‡]Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston IL 60208; and Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia PA 19104; e-mail: sandroni@kellogg.northwestern.edu.

meaning to the term *scientific*). He argued that science is not a collection of facts, but a collection of statements that can be falsified (i.e., conclusively rejected by the data). His leading example of a scientific statement was "All swans are white." This example shows the asymmetry between verification and refutation: no matter how many white swans are observed, one cannot be certain that the next one will be white, but the observation of a single nonwhite swan proves the statement to be false.

Popper's work is essentially conceptual and not descriptive. He argued that falsifiability is a criterion that should differentiate science from nonscience, but he did not claim that falsifiability is the only criterion used in practice to guide science, nor did he claim that theories produced by scientists are necessarily falsifiable. Indeed, several critics contend that the history of science contains instances that are inconsistent with Popper's criterion. Most notably, Thomas S. Kuhn (1962) argued that theories conclusively refuted by the data are not necessarily discarded in practice, that science sometimes makes use of nonfalsifiable theories, and that science often progresses by comparing competing theories and not by the falsification of theories.[1] In spite of these well-known limitations, falsifiability remains a central concept in the philosophy of science for several reasons. First, it presents one guiding principle on how science should be conducted: scientists should deliver falsifiable theories that can be tested empirically. (One example of falsifiability as a guide to research is the debate on whether general equilibrium theory is testable; see Andrés Carvajal, Indrajit Ray and Susan K. Snyder (2004) for a review article.) Moreover, falsifiability delivers criteria for what should be taught under the rubric of science. (One interesting application is the ruling by U.S. District Court Judge William Overton, largely based on falsifiability, against the teaching of intelligent design as science in Arkansas public schools, "The Arkansas Balanced Treatment Act" in *McLean* v. *Arkansas Board of Education*, Act 590 of the Acts of Arkansas of 1981.) Finally, falsifiability is an important requirement in the U.S. legal system's Daubert standard, which is designed to rule as inadmissible any testimony by expert witnesses that this standard evaluates as "junk science." (See the legal precedent set in 1993 by the Supreme Court, *Daubert* v. *Merrell Dow Pharmaceuticals*, 509 U.S. 579.)

Although falsifiability has been employed as a guiding principle in legal proceedings, economics, and science in general, it has not, to our knowledge, been formally analyzed to determine whether it can distinguish useful ideas from worthless ones in a full-fledged economic model in which agents may misreport what they know. An objective of this paper is to deliver such a model. Before continuing with a description of the model, we stress that what Popper means by falsifiability is the feasibility of *conclusive* empirical rejection. This is often regarded as too strong because it dis-

[1] See also Imre Lakatos' article in Lakatos and Alan Musgrave (1970) for an attempt to reconcile Popper's view on the logic of science with Kuhn's view on its history.

misses probabilistic statements that attach strictly positive probability to an event and its complement. Falsifiable probabilistic statements must attach zero probability to some event. Popper was aware of this objection. He wrote: "For although probability statements play such a vitally important rôle in empirical science, they turn out to be in principle *impervious* to strict *falsification*. Yet this very stumbling block will become a touchstone upon which to test my theory, in order to find out what it is worth." (See Popper (1968), pp. 133).

An adaptation of falsifiability designed to partially accommodate probabilities is provided by Antoine Cournot's (1843) principle, which states that unlikely events must be treated as impossible.[2] However, for reasons that will become clear at the end of this introduction, we refer to falsifiability in the strict Popperian sense.

We study a contracting problem between an expert and a tester. The expert, named Bob, announces a theory which is empirically tested by the tester named Alice. Like Popper, we assume that the main purpose of a theory is to make predictions. We define a theory as a mechanism that takes the available data as input and returns, as output, the probabilities of future outcomes. Before data are observed, Bob decides whether to announce a theory. If he does, he cannot revise his theory later. As data unfold, Alice tests Bob's theory according to the observed history.

Alice does not have a prior over the space of theories and is too ill-informed to formulate probabilities over the relevant stochastic process (i.e., she faces Knightian uncertainty). An expert could deliver these probabilities to her. If a theory is an accurate description of the data-generating process, then she benefits from the theory because it tells her the relevant odds (i.e., it replaces her uncertainty with common risk).[3] The difficulty is that Alice does not know if Bob is an informed expert who can deliver the data-generating process, or if he is an uninformed agent who knows nothing about the relevant process and who can deliver only theories unrelated to it.

We first assume that Alice takes Popper's methodology seriously and demands a falsifiable theory, i.e., a theory such that it predicts that some finite continuation of any finite history has zero probability. Alice pays Bob a small reward which gives utility $u > 0$, if he announces a falsifiable theory. In order to discourage Bob from delivering an arbitrary falsifiable theory, Alice proposes a contract that stipulates a

[2]Cournot (1843) was perhaps the first to relate the idea that unlikely events will not occur to the empirical meaning of probability. He wrote, "The physically impossible event is therefore the one that has infinitely small probability, and only this remark gives substance - objective and phenomenal value - to the theory of mathematical probability."

[3]Risk refers to the case where available information can be represented by a probability. Uncertainty refers to the case where the available information is too imprecise to be summarized by a probability. This distinction is traditionally attributed to Frank H. Knight (1921). However, Stephen F. LeRoy, and Larry D. Singell, Jr. (1987) argue that Knight did not have this distinction in mind.

penalty if Bob's theory is falsified in the future, i.e., if some history deemed impossible by the theory is eventually observed. This penalty gives Bob disutility $d > 0$. Bob receives no reward and no penalty if he does not announce any theory or if he announces a nonfalsifiable theory (in which case his utility is zero).

We now make a series of assumptions that are not meant to be realistic. Rather, they should be interpreted as an extreme case in which our result will be shown to hold, so that it will also be shown to hold under milder and more realistic conditions. These assumptions are: Alice eventually has an unbounded data set at her disposal and never stops testing Bob's theory unless it is rejected. Bob does not discount the future and so his contingent payoffs are $u - d$ if his theory is eventually rejected, and u if his theory is never falsified. Bob's liabilities are not limited and so the penalty d for having delivered a rejected theory can be made arbitrarily large, whereas the payoff u for announcing a falsifiable theory can be made arbitrarily small. Bob has no knowledge whatsoever of the data-generating process and so Bob, like Alice, also faces uncertainty and cannot determine the probability that any falsifiable theory will be rejected. Finally, Bob evaluates his prospects by the minimal expected utility he may obtain given any possible future realization of data.

Our last assumption is so extreme that it seems to settle the matter trivially. Assume that Bob announces any falsifiable theory f deterministically. Then many histories falsify f. Given any history that falsifies f, Bob's utility is $u - d$. Hence, under uncertainty, Bob's payoff for delivering any theory f deterministically is $u - d$. As long as the penalty for delivering a theory rejected by the data exceeds the reward for announcing a falsifiable theory, i.e., as long as $d > u$, Bob is better off not announcing any theory deterministically. It thus seems as if Alice can avoid the situation in which she receives a theory produced by an uninformed expert.

However, Bob still has one remaining recourse. He can randomize (only once) at period zero and select his falsifiable theory according to this randomization. This suffices. We show that no matter how large the penalty d, and no matter how low the reward u, there exists a way to strategically select falsifiable theories at random (i.e., according to specific odds that we describe explicitly) such that given *any* possible future realizations of the data, the expected utility of the random announcement of a theory exceeds the utility of not announcing any theory at all. At the heart of our argument is the demonstration that it is possible to produce falsifiable theories (at random) that are unlikely to be falsified, no matter how the data unfold in the future. Thus, Popper's strict falsification criterion (which requires a theory to assert that some events are impossible) cannot deter even the most ignorant expert because the feasibility of conclusive empirical rejection can be removed by strategic randomization.

There is a contrast between the case in which theories are exogenous (or delivered honestly) and the case in which theories may have been strategically produced. For an honest (exogenous) theory, falsifiability makes a fundamental conceptual distinction: falsifiable theories can be conclusively rejected while nonfalsifiable theories cannot. In contrast, when theories are produced by a potentially strategic expert, falsifiability does not impose significant constraints on uninformed experts and hence we cannot determine whether the expert is informed about the data-generating process.

The failure of falsifiability to deliver a useful criterion (when experts are strategic) motivates our analysis of the merits of verification and refutation as guiding principles for empirical research. This motivation can be understood as follows: Verification is typically a concern of those who view science as a source of explanations for observed phenomena. In contrast, refutation is typically a concern of those who view science as a source of testable hypotheses. Popper wanted to show a conceptual distinction between refutation and verification. He claimed that although past data could not deliver *conclusive* inferences about how the future will evolve, they could conclusively refute some hypothesis. This asymmetry drives the idea of the feasibility of conclusive rejection. However, in addition to the failure of Popper's falsifiability criterion to characterize the full range of actual science practice, our conceptual critique of Popper's falsifiability leads naturally to the question of whether there is any economically meaningful distinction between refutation and verification. We address this question in our general model where Bob is not restricted to announcing a falsifiable theory and there are many different ways in which Alice can evaluate Bob's theory.

So, now consider a general contract between Alice and Bob. If Bob accepts the contract, then he must deliver a theory before any data are observed. After the data is observed, Bob's theory is evaluated. In a *verification contract*, Alice pays Bob when his theory performs well in light of the evidence (e.g., the observed data are deemed consistent with Bob's theory), but Bob may pay Alice when he announces his theory. We make no restrictions on which data the contract can define as consistent with each theory. In a *refutation contract,* Bob pays Alice if his theory performs poorly in light of the evidence (e.g., the observed data are deemed inconsistent with Bob's theory), but Alice may pay Bob when he announces his theory. The falsification contract is therefore just a special refutation contract in which Alice pays nothing for nonfalsifiable theories, pays positive amounts for falsifiable theories, and Bob pays Alice if his theory is conclusively rejected. In other refutation contracts, Bob may pay Alice contingent on data that do not conclusively reject his theory.

We are interested in a screening contract which Bob, if informed, accepts, but if uninformed, does not accept. If informed, Bob faces risk and evaluates his prospects by his expected utility. If uninformed, Bob faces uncertainty and evaluates his prospects based on the minimum expected utility he may obtain contingent on any possible future realization of the data. A contract is *accepted by the informed expert* if Bob gets positive expected utility if he announces the actual data-generating process. A contract is *accepted by the uninformed expert* if Bob can select theories at random so that no matter which data are eventually observed, his expected utility will be positive.

We show that if the informed expert accepts any verification contract, then the uninformed expert also accepts this contract. This result has a basic implication: it is possible to produce theories (at random) that are likely to prove to be supported by the data, no matter how the data unfold. Hence, when experts are potentially strategic, both Popper's falsifiability and verification fail to provide useful criteria for the same reason: they cannot screen between informed and uninformed experts. In contrast, we show a refutation contract that can screen between informed and uninformed experts (i.e., informed experts accept the contract and uninformed experts do not accept it). This contract is based on an empirical test and a penalty for Bob if his theory is rejected by the test. The method for determining how to refute theories is novel and is not based on falsification or on any standard statistical test.

Popper was right to point out the asymmetry between verification and refutation. However, his argument is based on the idea that some empirical claims can be conclusively rejected, but not conclusively demonstrated. This asymmetry becomes immaterial when refutation is understood as (strict) falsification, and theories are produced by strategic experts. If refutation is understood more broadly (as it is in this paper), then, even if experts are strategic, the asymmetry can be established by proving the existence of a screening refutation contract and the nonexistence of a screening verification contract.

These two results deliver an original argument supporting the fundamental idea that refutation is a better maxim for empirical research than verification. In addition, the same asymmetry between verification and refutation still holds in some cases where permitted theories are restricted to be highly structured (e.g., if it is assumed that theories must be exchangeable). However, there exist restrictions on the set of permitted theories (e.g., the requirement that conditional probabilities must be fixed, independent of past evidence) which make possible the existence of both verification and refutation contracts that screen between informed and uninformed experts. In general, no matter which theories are allowed, if there exists a screening contract, then there exists a screening refutation contract.

Related literature

The idea that an ignorant agent can strategically avoid being rejected by an empirical test can be found in a number of papers (see Dean P. Foster and Rakesh V. Vohra (1998), Drew Fudenberg and David K. Levine (1999), Ehud Lehrer (2001), Alvaro Sandroni (2003), Sandroni, Rann Smorodinsky and Vohra (2003), Vladimir Vovk and Glenn Shafer (2005), and Wojciech Olszewski and Sandroni (2007, 2008)).

Some of these results are reviewed in Nicolò Cesa-Bianchi and Gábor Lugosi (2006).[4]

However, the idea that the concept of falsification can be analyzed as an empirical test, and that this test can be manipulated by ignorant experts, is novel. The classes of tests considered in the literature have not yet included the empirical test defined by falsification. In addition, the main issue analyzed in this paper - the merits of the concepts of verification and refutation for guiding research - is also not addressed in this literature.

Motivating Idea: Strategic Randomization

Consider a simple two-period model. In period one, a ball is drawn from an urn. The balls are of n possible colors. Alice does not know the composition of the urn. If informed, Bob has seen the composition of the urn. If uninformed, he has not.

Alice is willing to pay to become informed (i.e., to learn the composition of the urn), but she is concerned that Bob may be uninformed and would just give her an arbitrary distribution. Alice wants to discourage such a fraud. One difficulty is that if Bob tells Alice that any color is possible, then she cannot reject Bob's claim. So Alice takes Popper's advice and proposes a contract to Bob. If he accepts, he must deliver a falsifiable distribution (i.e., a probability measure over the n colors that assigns zero probability to at least one color) at period zero. That is, Bob must claim that at least one color is impossible. If none of the (allegedly) impossible colors is observed, then Bob's utility is $u > 0$. If an (allegedly) impossible color is observed, then Bob's utility is $u - d < 0$. If Bob does not accept Alice's contract, then his payoff is zero.

By requiring a falsifiable distribution, Alice may induce Bob to misrepresent what he knows (when all colors are possible). However, no matter what the composition of the urn, the probability of some color must be smaller than or equal to $1/n$. So, as long as

$$u \geq \frac{d}{n}, \tag{1}$$

Bob, whenever informed, is better off accepting Alice's contract and asserting that some color, among those least likely to occur, is impossible, than he is by not accepting Alice's contract. In addition, Bob has no incentive to misrepresent the relative odds of any colors other than the one he must claim to be impossible.

Now assume that Bob is uninformed. Then, he faces uncertainty and cannot determine the relevant odds of the colors. Let us assume that, under uncertainty, Bob determines the value of Alice's contract by the minimum expected utility he obtains, among all future realizations of the data.

[4] See Eddie Dekel and Yossi Feinberg (2006), Ehud Kalai, Lehrer and Smorodinsky (1999), Aldo Rustichini (1999), Lehrer and Eilon Solan (2003), Sergiu Hart and Andreu Mas-Colell (2001), Olszewski and Sandroni (2009a-b), Nabil I. Al-Najjar and Jonathan L. Weinstein (2008), Feinberg and Colin Steward (2008), and Lance J. Fortnow and Vohra (2009) for related results.

Assume that Bob announces any falsifiable distribution deterministically. For some outcomes, Bob is rejected and his utility is negative. Hence, under uncertainty, Bob's payoff is negative. It follows that Bob cannot accept Alice's contract and deliver a falsifiable distribution deterministically. This seems to suggest that Alice can screen between informed and uninformed experts at least when Bob is sufficiently averse to uncertainty. However, this is not true. Bob can produce a falsifiable distribution, at random, and obtain an expected positive payoff, no matter what the true composition of the urn might be.

Let p_i be a probability distribution that is falsified if and only if color i is realized (i.e., p_i assigns zero probability to i and positive probability to $j = 1, ..., n$, $j \neq i$). Assume that Bob selects each p_i, $i = 1, .., n$, with probability $1/n$. For *any* given color, Bob's realized probability measure is falsified with probability $1/n$. Hence, conditional on any composition of the urn, Bob's expected utility is nonnegative when (1) is satisfied. If Bob, when informed, accepts Alice's contract, then Bob, even if completely uninformed and averse to uncertainty, also accepts Alice's contract.

The argument above is simple, but it leads to an important implication. As pointed out earlier in the introduction, falsifiability is a criterion proposed by Popper to differentiate scientific ideas from nonscientific ideas. Namely, a falsifiable distribution can be conclusively rejected by the data and a nonfalsifiable one cannot. However, when theories are produced by experts who can misrepresent what they know, it is unclear whether falsifiability constitutes a useful criterion.

The argument in above was simple because Alice had only one data point at her disposal. In most relevant cases, however, Alice has many data points available to her. The puzzle of which criteria can be used (when experts are strategic) to distinguish useful from worthless theories then becomes far more interesting. We discuss the case of multiple data points in proposition 1 of section II (while the present example is offered as a simple illustration of proposition 1). However, before addressing the case of larger data sets, let us continue to consider the case of one data point, but with the additional requirement that Bob identify at least two colors that are impossible. The motivation here is to indicate (as we will show in detail in the formal parts of the paper) that additional criteria that may seem prima facie to be sensible for guiding research also suffer from the inability of screening between informed and uninformed experts. Let us say that Bob must assert that j colors are impossible. Then, an informed expert always accepts Alice's contract if (1) is satisfied when $1/n$ is replaced with j/n. However, under this assumption, we can show that an uninformed expert can randomize and obtain a positive expected payoff no matter what the composition of the urn is. Hence, the additional requirements for falsification may not suffice for screening between informed and uninformed experts.

I. Basic Definitions

We now consider a model with many periods so that Alice will eventually have a large data set available to her. Each period, one outcome, out of a finite set S with n elements, is observed. Let $\Delta(S)$ denote the set of probability measures on S. An element of $\Delta(S)$ is called a *distribution over outcomes*. Let S^t denote the Cartesian product of t copies of S, and let $\overline{S} = \bigcup_{t \geq 0} S^t$ be the set of all finite histories.[5] We also define $\Omega = S^{\infty}$ as the set of *paths*, i.e., infinite histories, and we define $\Delta(\Omega)$ as the set of probability measures over Ω.[6]

Any function $f : \overline{S} \longrightarrow \Delta(S)$ that maps finite histories into distributions over outcomes can be interpreted as follows: f takes data (outcomes up to a given period) as input and returns a probabilistic forecast for the following period as output. To simplify the language, any such function f is called a *theory*. Thus, a theory is defined by its predictions. Let $\mathcal{T}$ be the set of all theories.

Any theory $f \in \mathcal{T}$ defines a probability measure P_f. The probability of each finite history $(s^1, ..., s^m) \in S^m$ can be computed as follows: Given a finite history $\overline{s} \in \overline{S}$ and an outcome $s \in S$, let the probability of s conditional on $\overline{s}$ be denoted by $f(\overline{s})[s]$. Then the probability P_f of $(s^1, ..., s^m)$ is equal to a product of probabilities

$$P_f(s^1, ..., s^m) = f(\varnothing)[s^1] \cdot \prod_{k=2}^{m} f\left(s^1, ..., s^{k-1}\right)[s^k]. \tag{2}$$

Definition 1. *A theory f is falsifiable if every finite history $(s^1, ..., s^t) \in S^t$ has an extension $(s^1, ..., s^t, s^{t+1}, ..., s^m)$ such that*

$$P_f(s^1, ..., s^m) = 0. \tag{3}$$

A theory is falsifiable if, after any finite history, there is a finite continuation history that the theory deems impossible.[7] Let $\mathcal{F} \subset \mathcal{T}$ be the set of falsifiable theories. Given $f \in \mathcal{F}$, let R_f be the set of all finite histories to which P_f assigns zero probability. So, R_f is the set of all finite histories that contradict (or, equivalently, falsify) the theory $f \in \mathcal{F}$.

[5] By convention, $S^0 = \{\varnothing\}$.

[6] We need a $\sigma-$algebra on which probability measures are defined. Let a *cylinder* with base on $(s^1, ..., s^m)$ be the set of all paths such that the first m elements are $(s^1, ..., s^m)$. We endow Ω with the smallest $\sigma-$algebra that contains all such cylinders and with the product topology (the topology that comprises unions of cylinders). We also endow $\Delta(\Omega)$ with the weak$-*$ topology.

[7] One could say that conclusive rejection only requires some finite history to be impossible. Popper makes no comment on whether theories should remain falsifiable after data is observed. In proposition 1, we show that strategic experts can avoid rejection even if theories are required (by definition 1) to remain permanently falsifiable. If we adopt a transient (and hence weaker) concept of falsifiability (i.e., theories need to falsified just once), then our manipulability result still holds.

Assume that Alice demands a falsifiable theory. Alice pays for the theory, but if it is rejected (i.e., if Bob announces theory $f \in \mathcal{F}$ and data in R_f is observed), then Bob receives a large penalty that gives a disutility greater than the utility of the payment. The question is whether the falsifiability requirement dissuades an uninformed expert (who does not know the data-generating process) from announcing a theory.

We show next that an uninformed expert can strategically produce falsifiable theories, at random, with odds designed so that, with arbitrarily high probability, the realized falsifiable theory will not be falsified, regardless of which data are eventually observed. Hence, the feasibility of falsification is virtually eliminated by strategic randomization.

II. Falsification and Strategic Randomization

In this section, we maintain the basic infinite horizon model in which Alice adopts Popper's method and demands a falsifiable theory $f \in \mathcal{F}$. As an incentive, Alice pays Bob a (small) amount of money (which gives Bob utility $u > 0$) if Bob delivers a falsifiable theory. However, if a finite history in R_f is observed, then Bob's theory is falsified and he pays a penalty which gives him disutility $d > 0$. Liabilities are not limited and so d can be arbitrarily large.

Bob does not have to deliver a theory, but if he accepts Alice's conditions, he must deliver a falsifiable theory before any data are observed. We assume that Bob does not discount the future (although our result still holds if he does). So Bob's contingent payoffs are $u > 0$ if his theory is never falsified, and $u - d$ if his theory is contradicted at some time t. Formally, consider the contract in which only falsifiable theories can be delivered, and delivering a theory which is (later) contradicted by data is punished. If Bob accepts the contract, he announces a theory $f \in \mathcal{F}$. If a path $s = (s^1, s^2, ...) \in \Omega$ is observed, Bob's contingent net payoff, at period zero, is

$$U(f, s) = \begin{cases} u - d & \text{if, for some period } t,\ (s^1, ..., s^t) \in R_f; \\ u & \text{otherwise.} \end{cases} \tag{4}$$

We call this contract the *falsification* contract. If Bob does not accept the falsification contract, then no theory is announced and his payoff is zero.

We assume that Bob is utterly ignorant about the relevant probabilities. So, like Alice, Bob cannot determine the odds according to which any given theory will be falsified. However, Bob can select his theory randomly according to a probability measure $\zeta \in \Delta(\mathcal{F})$. Given that Bob randomizes only once (at period zero), Alice cannot tell whether the theory she receives was produced deterministically or selected randomly. Finally, we assume that when Bob faces uncertainty, he evaluates his prospects based on the future data path $s \in \Omega$ that gives him minimal expected utility. Formally, Bob's payoff is

$$V(\zeta) \;=\; \inf_{s \,\in\, \Omega} E^{\zeta} U(f, s), \tag{5}$$

where E^{ζ} is the expectation operator associated with Bob's randomization device $\zeta \in \Delta(\mathcal{F})$.

If Bob announces *any* theory $f \in \mathcal{F}$ deterministically, then his payoff is $u - d$ because for every theory $f \in \mathcal{F}$, there are many paths (i.e., those in R_f) at which f will be falsified. Hence, as long as the punishment d is greater than the reward u, Bob finds that announcing any falsifiable theory deterministically is strictly worse than not announcing any theory at all. Formally, as long as $d > u$,

$$\inf_{s \,\in\, \Omega} U(f, s) < 0 \text{ for every } f \in \mathcal{F}. \tag{6}$$

However, Bob can randomize and, as Proposition 1 shows, randomization alters Bob's prospects completely.

Proposition 1. *For any payoffs $u > 0$ and $d > 0$, and for any $\rho > 0$ smaller than u, there exists a randomization device $\bar{\zeta} \in \Delta(\mathcal{F})$ such that, given any future realization of the data, Bob's expected utility is strictly positive and bounded away from zero. That is,*

$$V(\bar{\zeta}) \; > u - \rho > 0. \tag{7}$$

Proposition 1 shows that no matter how small the rewards for delivering a falsifiable theory, no matter how large the penalties for having a theory falsified, and no matter how much data Alice might have at her disposal, Bob is strictly better off by accepting her contract and producing a theory at random. Even if Alice demands a falsifiable theory, she will not dissuade the most ignorant expert from delivering a fraudulent theory to her. This holds even in the extreme case that this ignorant expert evaluates his prospects based on the minimum expected utility he receives among *all* possible realizations of the data.

The striking contrast between the case of an honestly revealed theory and the case of strategically produced theories conveys the first part of our argument. Falsifiability can be, and often is, used as a relevant criterion. In some cases, this criterion may seem intuitively weak (e.g., when there are too many possible outcomes and only a single one must be ruled out). In other cases, this criterion may seem stronger (e.g., when a future outcome must be ruled out in every period). Still, whether intuitively weak or strong, if an expert is honest and wants his theory rejected (when false), then falsifiability is a useful criterion because only falsifiable theories can be conclusively rejected.

In contrast, the requirement that theories must be falsifiable becomes near irrelevant when theories are produced by a potentially strategic expert, because if we only require theories to be falsifiable, then we cannot discredit experts who are completely uninformed about the data-generating process. This result casts doubt on the idea that falsifiability can demarcate legitimate theories from worthless theories. As long as theories are produced by experts capable of strategic randomization, the falsification criterion cannot screen informed from uninformed experts.

A. *Intuition that underlies Proposition 1*

Given a random generator of falsifiable theories $\zeta \in \Delta(\mathcal{F})$ and a path $s = (s^1, ..., s^t, ...) \in \Omega$, the odds that the selected theory f will be contradicted at some point in the future are

$$p_\zeta(s) = \zeta\left\{f \in \mathcal{F} \mid \text{there exists } t \text{ such that } (s^1, ..., s^t) \in R_f \right\}. \tag{8}$$

The key argument is that for every $\varepsilon > 0$ there exists a random generator of falsifiable theories $\bar{\zeta} \in \Delta(\mathcal{F})$ such that the odds of selecting a theory that will be falsified is smaller than ε, *for every path* $s \in \Omega$. That is,

$$p_{\bar{\zeta}}(s) \leq \varepsilon \text{ for every path } s \in \Omega. \tag{9}$$

No matter which data are realized in the future, a falsifiable theory selected by $\bar{\zeta}$ will not be falsified, with arbitrarily high probability. Hence, by randomizing according to specific probabilities, Bob is near certain that his theory will not be falsified.

To construct $\bar{\zeta}$, we first consider an increasing sequence of natural numbers Z_t, $t = 1, 2....$ Let $X_t = S^{(Z_{t+1} - Z_t)}$ be the set of outcome sequences of length $(Z_{t+1} - Z_t)$. Given any sequence $x = (x_t)_{t=1}^{\infty}$, where each $x_t \in X_t$ consists of $(Z_{t+1} - Z_t)$ outcomes, we define a falsifiable theory f_x which is falsified if and only if x_t occurs between periods Z_t and Z_{t+1}. The random generator of theories $\bar{\zeta}$ is then defined as follows: an x_t, $t = 1, 2, ...$, is chosen from a uniform probability distribution over X_t, and the sequence $x = (x_t)_{t=1}^{\infty}$ (thus chosen) determines the theory f_x that is announced.

If Z_t grows sufficiently fast (and if z_1 is sufficiently large), the chance that x_t will be realized at least once is small. Any path s can be written in the form $s = (y_1, ..., y_t, ...)$, where y_t consists of $(Z_{t+1} - Z_t)$ outcomes. So the announced theory f_x is contradicted along s if and only if $x_t = y_t$ for some period t. By construction, this is an unlikely event.

B. *Extension of Proposition 1 to Multiple Testers*

In proposition 1, Alice is the only tester. More generally, there could be multiple testers $l = 1, ...L$. Each tester eventually has an arbitrarily large stream of data at her disposal. Let $s(l) = (s^1(l), s^2(l), ...) \in \Omega$ be a data path obtained by tester l. Let

$(s(1), ..., s(L)) \in \Omega^L$ be a profile of data paths obtained by the L testers. Now assume that Bob must deliver a single falsifiable theory f at period zero to all testers. If, at any period t, Bob's theory f is falsified by the $t-$history $(s^1(l), s^2(l), ...s^t(l))$ obtained by *any* tester l, then Bob's theory is rejected.

Proposition 1 extends to the case of multiple testers. As in proposition 1, the central argument is the demonstration that it is possible to produce falsifiable theories (with a single, properly chosen randomization device) that are unlikely to be falsified by any data of the L testers.

Fix $\varepsilon > 0$. Let $\bar{\zeta}^L \in \Delta(\mathcal{F})$ be the random generator of falsifiable theories such that (9) holds with ε replaced by ε/L. Given any data path profile $(s(1), ..., s(L)) \in \Omega^L$, the odds that a theory f (selected by $\bar{\zeta}^L$) is rejected by any tester is $p_{\bar{\zeta}^L}(s(1), ..., s(L)) =$

$$\bar{\zeta}^L \left\{ f \in \mathcal{F} \mid \text{for some period } t \text{ and tester } l,\ (s^1(l), ..., s^t(l)) \in R_f \ \right\}. \tag{10}$$

The event in which theory f is falsified by some tester l is the union of the events in which f is falsified by tester $l = 1, ..., L$. Hence, for *any* data path profile $(s(1), ..., s(L)) \in \Omega^L$,

$$p_{\bar{\zeta}^L}(s(1), ..., s(L)) \leq \varepsilon. \tag{11}$$

It is thus unlikely that the (single) theory produced by $\bar{\zeta}^L$ will be rejected by the data obtained by any of the L testers at any point in time, no matter how their data unfold in the future.

C. *Popper's Falsifiability Revisited*

Science is sometimes identified with a collection of ideas that have been partially confirmed by facts, and are capable of providing some explanation for observed phenomena. As mentioned in the introduction, Popper contrasted this popular view of science with the notion that science must only propose conjectures that can be empirically refuted. Hence, a central question for Popper is whether there exists a meaningful distinction between verification and refutation. He answered this question affirmatively by revisiting the classic problem of induction and pointing out that while, under suitable conditions, one can use past data to argue that some outcome is likely to occur, it is difficult to argue conclusively from past observations that it must occur. In contrast, some claims can be conclusively refuted with a single observation. Hence, in Popper's work, the fundamental asymmetry between verification and refutation lies in the concept of falsifiability: the feasibility of *conclusive* rejection.

Popper's ideas have been intensively debated. One important critique is that falsifiability rules out probabilistic statements that attach strictly positive probability everywhere (i.e., all finite events). This difficulty with falsification can be mitigated by the Cournot principle. If Bob's theory is not falsifiable, then he can still satisfy Popper's criterion by reporting a modified theory that attaches zero probability to some unlikely events. Falsification can therefore be understood as giving Bob some flexibility in the determination of which data can reject his theory.

If Bob is an honest expert who reports sufficiently implausible events to merit rejection of his premises, then his theory may be refuted. However, this intuition does not extend to the case of strategic experts who may deliver falsifiable theories that claim that only few events (out of all possible events) are impossible and then, by properly randomizing among these falsifiable theories, virtually eliminate the odds of rejection.

The failure of Popper's falsifiability to discredit false, but strategic, experts motivates an analysis of alternative criteria for guiding scientific inquiry. One natural direction is to allow any probabilistic statements, but to give the tester more discretion regarding how to evaluate each theory. In the next section, we formally define the broader concepts of verification and refutation as properties of the empirical methods that Alice may use. The central question is whether there exists an economically meaningful distinction between the two methods.

III. Verification and Refutation Contracts

In this section, we consider general contracts. The section is organized as follows: We first define a general contract, and two important subclasses of the class of all contracts: verification and refutation contracts. We show that verification contracts cannot screen informed and uninformed expects, and give experts very poor incentives for information acquisition. We discuss this, perhaps surprising, result in section II B. Next, we show that in contrast some refutation contracts can screen informed and uninformed expects, and discuss this result in sections II D and E. One important conclusion coming out of this discussion is that the asymmetry between verification and refutation holds true not only when theories are unrestricted, but also for some highly structured theories, such as exchangeable processes. Finally, we show that restricting attention to refutation contracts, as opposed to studying general contracts, involves no loss of generality in terms of the possibility of screening experts.

Bob decides whether to accept a contract at period zero. If Bob does not accept a contract, then he does not deliver a theory and his payoff is zero. If Bob accepts a contract, then he delivers a theory $f \in \mathcal{T}$, which is now not required to be falsifiable, to Alice at period zero (before any data are observed). An initial transfer may occur at period zero (after the theory is announced). This transfer gives utility $u(f, \varnothing)$ to Bob. At period t, if data $s_t \in \overline{S}$, $s_t = (s^1, ..., s^t)$, are observed, then a new transfer may occur. Bob's payoff, evaluated at period zero, for this contingent transfer is $u(f, s_t)$. So, given a path $s \in \Omega$, Bob's contingent payoff at period zero is

$$U(f,s) = u(f,\varnothing) + \sum_{t=1}^{\infty} u(f,s_t), \text{ where } s = (s^1, ..., s^t, ...). \tag{12}$$

The utility functions u and U are assumed to be bounded. Therefore, Bob cannot receive unboundedly large contingent payoffs (either positive or negative) in any single period or in total. This assumption is maintained throughout the paper unless otherwise specified, but no other restrictions, such as continuity, are imposed on the utilities associated with Alice's contract.[8]

In this model, the payoffs are defined very generally. Hence, a variety of ways to evaluate theories can be accommodated by this framework. Alice may judge some theories to be prima facie more useful than others; in addition, Bob's initial payoff $u(f,\varnothing)$, which is contingent only on his theory, may partially reflect judgments about theories without data. In addition, the contract specifies payoffs $u(f,s_t)$ contingent on the announced theory f and the observed data s_t. Some contracts may specify which evidence disproves each theory and prescribe a negative payoff if the theory is disproved, while other contracts may specify which evidence confirms each theory and prescribe a single positive payoff if the theory is confirmed. Still other contracts may never disprove or confirm theories. Instead, they specify positive payoffs when the weight of the evidence leans in favor of the theory and negative payoffs when the weight of the evidence leans against the theory. In general, contracts in our model may not rely on any notion of whether the evidence does or does not supports each theory, but rather on how helpful for Alice the theory is, contingent on observed evidence. The key assumption is that experts receive different payoffs depending on the announced theory and how it performs in light of future evidence, but the way in which theories are evaluated is quite flexible and intentionally left to be specified by the contract.

We now consider two fundamentally different types of contracts. If $u(f,s_t) \geq 0$ for every $s_t \in \bar{S}$, $s_t \neq \varnothing$, then the contract is said to be a *verification contract*. If $u(f,s_t) \leq 0$ for every $s_t \in \bar{S}$, $s_t \neq \varnothing$, then the contract is said to be a *refutation contract*. So, although Bob may receive a (positive, negative, or zero) payoff at period zero when the theory is delivered, the distinction between verification contracts and refutation contracts depends only on the payoffs after the data are observed. Verification contracts are those in which, after the data are revealed, Bob receives either no payoff or a positive payoff, contingent on the performance of the theory vis-a-vis the data. Refutation contracts are those in which, after the data are revealed, Bob receives either no payoff or negative payoffs contingent on how his theory performs vis-a-vis the data. The terminology reflects the idea that in a verification contract

[8] Naturally, we must impose standard measurability conditions to study Bob's expected payoffs.

Bob is paid when his theory performs well, while in a refutation contract it is Bob who pays when his theory performs poorly.

The distinction between falsification and refutation can be understood as follows. Falsifiability is a specific restriction on the set of allowed theories, coupled with an indication of how to reject each permitted theory. So, of all possible theories only the subset of falsifiable theories $\mathcal{F} \subset \mathcal{T}$ is allowed. Bob receives utility u for announcing any falsifiable theory. Each falsifiable theory is rejected when an event deemed impossible by the theory is observed. Bob receives disutility d if his theory is falsified. In contrast, refutation (and also of verification) contracts need not, by themselves, restrict the set of allowed theories. Unlike falsification, a refutation contract does not necessarily dismiss a nonfalsifiable theory, or indeed any theory, before data are observed. It may still pay for the nonfalsifiable theory and punish Bob later if his theory does not perform adequately vis-a-vis the data in the sense defined by the refutation contract.

However, it is possible to combine the idea of verification contracts and refutation contracts with restrictions on the set of allowed theories. Any contract could restrict the set of allowed theories to some set (or class) $\mathcal{A} \subset \mathcal{T}$. In this paper, the set of allowed theories take this form: $f \in \mathcal{A}$ is permitted if and only if $P_f \in \Theta \subseteq \Delta(\Omega)$, where Θ is a set of stochastic processes. So, we may refer to Θ as a restriction on permitted theories. In the case that $\Theta = \Delta(\Omega)$, we say that theories are not restricted.

Given a theory $f \in \mathcal{T}$ and a probability measure $P \in \Delta(\Omega)$, we define

$$\bar{U}^P(f) = E^P U(f, s), \tag{13}$$

where E^P is the expectation operator associated with P. If Bob is informed at period zero and announces a theory f, then his expected payoff is $\bar{U}(f) = \bar{U}^{P_f}(f)$. We say that an *informed expert accepts the contract* if, for *all* allowed theories $f \in \mathcal{A}$,

$$\bar{U}(f) > 0. \tag{14}$$

So, if Bob is informed, he knows the odds of future events (i.e., the real process that will generate s), and (strictly) prefers to announce what he knows rather than to refuse the contract.[9]

Now assume that Bob faces uncertainty and does not know anything about the data-generating process. In this case, we say that Bob is uninformed. As in the case of the falsification contract, Bob can select his theory by randomizing once (at period zero) according to a random generator of theories $\zeta \in \Delta(\mathcal{A})$. If uninformed, Bob

[9] It is convenient to assume that, he is indifferent between accepting and rejecting the contract, the expert (whether informed or not) rejects the contract. Similar results can be obtained (from the authors upon request) under the opposite convention in which experts, when indifferent, accept the contract.

evaluates his prospects based on the process $P \in \Theta$ that gives him minimal expected utility. Formally, Bob's payoff is

$$V(\zeta) = \inf_{P \in \Theta} E^P E^\zeta U(f, s), \tag{15}$$

where E^ζ is the expectation operator associated with $\zeta \in \Delta(\mathcal{A})$. In the case of unrestricted theories, Bob's payoff is simply

$$V(\zeta) = \inf_{s \in \Omega} E^\zeta U(f, s). \tag{16}$$

We say that the *uninformed expert accepts the contract* if there exists a random generator of theories $\bar{\zeta} \in \Delta(\mathcal{A})$ such that

$$V(\bar{\zeta}) > 0. \tag{17}$$

So, an uninformed expert accepts the contract if he can randomly select theories such that, contingent on any possible realization of the data, his expected payoff with the contract is positive.

A contract *screens between informed and uninformed experts* if the informed expert accepts the contract, but the uninformed expert does not accept it.

Under uncertainty, Bob's payoff is given by (15). His payoff is determined under the worst-case scenario, i.e., the path that gives him minimum expected utility with his random strategy ζ. The worst-case scenario is computed ex-ante (i.e., before Bob randomizes) and not ex-post (after Bob selects his theory). Indeed, if Bob were to compute his payoff under the worst-case scenario after he randomizes, then he would not take any contract such that, for every theory f, there is some path s such that the payoff $U(f, s)$ is negative.

While there may be more than one way of providing a rationale for the payoffs in (15) and (16), in the case of unrestricted theories, choice-theoretic foundations can be provided by the axiomatic framework of Itzhak Gilboa and David Schmeidler (1989). In the case of restricted theories, Olszewski (2007), David S. Ahn (2008) and Thibault Gajdos, Takashi Hayashi, Jean-Marc Tallon and Jean-Christophe Vergnaud (2008) developed models in which the subjective range of probabilities is a subset of the set of all objectively possible probabilities.

In Gilboa and Schmeidler (1989) maxmin expected utility model, Bob chooses as if he had preferences represented by

$$\inf_{P \in \tilde{\Theta}} E^P E^\zeta U(f, s), \tag{18}$$

where $\tilde{\Theta} \subseteq \Delta(\Omega)$ is a closed and convex set of probability measures. So, (16) is equivalent to the special, and polar case of (18) in which $\tilde{\Theta} = \Delta(\Omega)$ is the set of all measures. Hence, if the uninformed expert with payoffs (16) accepts a contract, then he accepts the contract in any maxmin expected utility model. But if the uninformed expert rejects the contract, he rejects the contract only in some such models.

The critical axiom in Gilboa and Schmeidler is *uncertainty aversion*: if Bob is indifferent between acts f and g, then he prefers any mixture of the two acts to both f and g. We exploit this property in our results. The uninformed expert will accept some contracts because a randomization over theories gives a positive payoff, even though any single theory gives a negative payoff. We now describe our results.

Proposition 2. *Assume that theories are unrestricted.*[10] *If an informed expert accepts a verification contract, then an uninformed expert also accepts this verification contract.*

Proposition 2 shows that verification contracts cannot screen between informed and uninformed experts. This result shows a fundamental limitation on verification as a guiding principle for empirical analysis. No verification contract can mitigate Alice's adverse selection problem.

Now consider the following moral hazard problem: Bob is uninformed, but can become informed before any data are revealed, if he acquires sufficient information that allows him to formulate probabilities (i.e., to transform his uncertainty into risk). The cost (i.e., his disutility) of acquiring this information is $c > 0$. When Bob decides whether to become informed, he does not know the data-generating process and he does not have a prior over the space of data-generating processes (otherwise he would use this prior to assess the odds of the data). Hence, Bob makes a decision under uncertainty. As before, we assume that Bob evaluates his prospects based on the minimum expected utility he may obtain. Therefore, Bob's net value for becoming informed is

$$V(I, c) = \inf_{f \,\in\, \mathcal{A}} \bar{U}^{P_f}(f) - c. \tag{19}$$

That is, $V(I, c)$ is Bob's smallest expected utility when informed, minus the cost c of becoming informed. On the other hand, if Bob remains uninformed and produces a theory according to ζ, then his payoff is still $V(\zeta)$.

We say that the expert *prefers not to become informed* if there exists a random generator of theories $\bar{\zeta}$ such that $V(\bar{\zeta}) > V(I, c)$. In that case, Bob prefers to announce theories at random (selected by $\bar{\zeta}$) to becoming informed at cost c.

Proposition 3. *Assume that theories are unrestricted.*[11] *Consider any verification contract and any positive (possibly arbitrarily small) cost $c > 0$ of acquiring information. Then, the expert prefers not to become informed.*

[10] More generally, theories may be restricted by any closed and convex set $\Theta \subseteq \Delta(\Omega)$.

[11] As in proposition 2, this result also holds with restricted theories where Θ is closed and convex.

Proposition 3 shows that no matter how low the cost of acquiring information is, the expert prefers not to acquire it. Instead, the expert chooses to present a randomly selected theory.

To underscore these results, assume that for every theory f, a set of finite histories A_f are deemed consistent with f. Also assume that A_f has probability $1-\varepsilon$, $\varepsilon > 0$, under theory f (i.e., $P_f(A_f) = 1-\varepsilon$).[12] Hence, if the announced theory is indeed the data-generating process, then the data are likely to be deemed consistent with the theory, but no other restrictions are placed on which histories are deemed consistent with each theory.

Now consider the contract in which Bob receives zero payoff in period zero, when the theory is announced, but receives payoff 1 whenever the observed history is deemed consistent with Bob's theory. (Formally, $u(f, \varnothing) = 0$, $u(f, s_t) = 1$ if $s_t \in A_f$, and $u(f, s_t) = 0$ if $s_t \notin A_f$.) At period zero, Bob's net value of becoming informed, $V(I, c)$, is $1-\varepsilon-c$ because, if informed, Bob gets payoff 1 with probability $1-\varepsilon$. By proposition 3, $V(\bar{\zeta})$ is higher than $1-\varepsilon-c$ for some $\bar{\zeta} \in \Delta(\mathcal{T})$. This implies that no matter how the data unfold, a theory selected by $\bar{\zeta}$ will be deemed consistent with the data, with probability sufficiently close to $1-\varepsilon$. Formally, there exists $\bar{\zeta} \in \Delta(\mathcal{T})$ such that for all $s \in \Omega$,

$$\bar{\zeta}\{f \in \mathcal{T} \mid s_t \in A_f \text{ for some } t\} > 1-\varepsilon-c. \tag{20}$$

Thus, no knowledge over the data-generating process is necessary to produce theories that will, in the future, prove to be supported by the data. The widespread practice of supplying theories that account for the facts is vulnerable to the usual reproach centered at the existence of alternative theories that also explain the data. However, as we have demonstrated, even *without knowing the data, uninformed "experts" can fabricate theories that will be supported by the data.* This result makes the standard critique salient.

A. *Intuition underlying Propositions 2 and 3*

Although the main contribution of this paper is conceptual, we wish to point out that the proofs of propositions 2 and 3 use novel arguments, but rely heavily on a combination of analytical techniques developed in the strategic experts literature (particularly in Olszewski and Sandroni (2007) and (2009b)). The former paper considers a class of contracts that cannot screen informed and uninformed experts. These contacts may be verification contracts, refutation contracts or none of these two types (the verification/refutation distinction is not made and addressed in these

[12] For example, if Bob's theory asserts that 1 has probability $p \in [0, 1]$ in all periods, then an acceptance set could comprise all histories in which the relative frequency of 1 is between $p-\delta$ and $p+\delta$, after sufficiently many periods.

papers). The latter paper focus on test-based contracts and contains a result which is subsumed by proposition 2.

The proofs of propositions 2 and 3 are similar, and so we focus on the intuition for proposition 3. Consider the zero-sum game between Nature and the expert such that Nature's pure strategy is a path $s \in \Omega$ and the expert's pure strategy is a theory $f \in \mathcal{T}$. The expert's payoff is $U(s, f)$. For every mixed strategy of Nature P, there exists a strategy for the expert (to announce a theory f such that $P_f = P$) that gives the expert an expected payoff of at least $V(I, c) + c$. So, if the conditions of Ky Fan's (1953) minmax theorem are satisfied, there is a mixed strategy for the expert, $\bar{\zeta}$, that gives him an expected payoff higher than $V(I, c)$, no matter which path $s \in \Omega$ Nature selects.

A key condition in Fan's minmax theorem is the lower semi-continuity of the expert's payoff. In refutation contracts, the expert's payoff is not necessarily lower semi-continuous; in contrast, the positive payoffs in verification contracts suffice to make $U(f, s)$ a lower semi-continuous function of s. By definition, $U(f, s)$ is a lower semi-continuous function of $s = (s^1, ..., s^t, ...)$ if for every path s' which coincides with s on a sufficiently long initial history (i.e., $s_t = (s^1, ..., s^t) = s'_t$ if t is sufficiently large), $U(f, s')$ is not much *lower* than $U(f, s)$. If $U(f, s)$ is bounded, then, when t is sufficiently large, $U(f, s)$ is approximately equal to the sum of the utilities u received up to period t. Since these utilities depend only upon the history up to period t and $s_t = s'_t$, the sum of the utilities received up to period t is the same on paths s and s'. The payoffs received after period t on path s' can be significant, so $U(f, s')$ may not be approximately equal to the sum of the utilities received up to period t. In a verification contract, however, the payoffs received after period t are positive, so they can make $U(f, s')$ only *higher*.

The assumption of lower semi-continuity plays a role in Fan's minmax theorem that is similar to the role of all sorts of continuity conditions on payoff functions for the existence of Nash equilibrium in games. (See Philip J. Reny (1999), among many others, for existence of Nash equilibrium results in some discontinuous games.) Similarly, the role of the assumption that the set of allowed theories is unrestricted is that the strategy space for the expert is compact and convex. So, as mentioned in footnotes 12 and 13, propositions 2 and 3 still hold even if the expert is not allowed to announce any theory, but is instead restricted to announcing a theory f such that P_f belongs to some closed and convex Θ. We exploit this extension of propositions 2 and 3 in the next section.

B. *Verification Contracts, Experimentation, and Structured Theories*

Consider first the case in which Alice can replicate identical experiments. So, in each period Alice conducts the same experiment, and in each period she obtains an outcome that is assumed to be generated with the same probability distribution. In this environment, Alice can eventually uncover, to a certain degree of precision, all relevant probabilities from the data (even without any need for Bob). The relative frequencies of outcomes will reveal the probabilities of outcomes. Now, consider a verification contract in which Bob must deliver the fixed probability of each outcome at period zero, and gets paid if and only if at a sufficiently distant period, the relative frequencies of outcomes are sufficiently close to the announced probabilities. Bob is then likely to get paid if and only if he announces nearly correct probabilities at period 0. A formal verification contract, based on a proper match between announced probabilities and observed frequencies, can be written such that the informed expert accepts the contract and the uninformed expert rejects it. Analogously, a screening refutation contract can also be constructed if multiple identical experiments can be conducted.

It follows that screening contracts exist if Alice can replicate identical experiments, or the allowed stochastic processes are independent and identically distributed (*i.i.d.*). However, in several relevant situations, it may be difficult to produce identical experiments, or it is unclear whether relevant variables must follow an *i.i.d.* process. Examples include weather forecasting, or predicting basic economic variables such as inflation and unemployment. Therefore, let us consider instead the still strong, but less demanding restriction to exchangeable processes. In this case, the probabilities of data sequences must not depend on the order of the entries of the data.[13]

Exchangeable processes need not be *i.i.d.*, but they are still highly structured, stationary processes (e.g., a Polya's urn) which are amenable to specific types of empirical investigation.[14] For example, if the process is exchangeable, then relative frequencies must converge. (See David M. Kreps (1988) for a more detailed discussion on the central role of exchangeable processes in economics.)

Bruno De Finetti (1937) celebrated result shows that exchangeable processes are mixtures of independent, identically distributed processes and form a closed and convex set. It is for this reason that propositions 2 and 3 hold even if the expert is restricted to announcing an exchangeable process at period zero. The substantive

[13]Formally, a exchangeable theory can be defined as follows : Let a one-to-one mapping $\pi : N \rightarrow N$ from the set of natural numbers onto the set of natural numbers be called a *permutation.* Given a permutation π, let $Y^{\pi} : \Omega \rightarrow \Omega$ be defined by

$$Y^{\pi}(s^1, s^2, ...) = (s^{\pi(1)}, s^{\pi(2)}, ...).$$

A theory f is *exchangeable* if given any permutation π,

$$P_f(A) = P_f(Y^{\pi}(A))$$

for any measurable event A.

[14]Note that exchangeable processes may not be falsifiable, and we do not restrict them to be falsifiable.

point here is that verification contracts may remain incapable of screening between informed and uninformed experts even in some cases where Bob is required to deliver highly structured theories such as an exchangeable process, or a mixture of other well-known processes like Markov processes. (See Persi Diaconis and David Freedman (1987) for a series of De Finetti's type results.)

We conclude this section with some simple examples of a class of processes for which there exist both verification and refutation screening contracts, and a class of processes for which no screening contract of either type exists. Assume that there are only two possible outcomes, 0 and 1, in each period, and theories are restricted to predict either 0 in every period or 1 in every period (with certainty in both cases). The set of allowed theories is then *not* closed and convex. In addition, one data point perfectly identifies the underlying process. Thus, both screening verification and refutation contracts can be written. For a verification contract, assume that Bob pays 0.5 in period 0, and gets 1 if and only if his forecast in period 1 is correct. If informed, Bob's total payoff is 0.5. If uninformed, Bob's total payoff is -0.5. Hence, the informed expert accepts the contract while the uninformed expert does not.

Mixtures of these two processes are theories such that any probability of outcome 1 in the first period is allowed, combined with the condition that 1 (or 0) must occur thereafter if and only if 1 (or 0) has been observed in the previous periods. This is a closed and convex class of processes. If Bob's theory (and the data-generating process) are restricted to this class, then the data will not perfectly reveal the underlying process. In addition, Alice will effectively have only one data point at her disposal (either 0 or 1 in the first period) to evaluate Bob's probability of 1 in the first period. In this case, no screening contracts (verification, refutation, or other) exist.

Finally, one can argue that in practice the forecasters may often not know in advance how they will be rewarded or penalized for their forecast contingent on the data that will be observed in the future. The assumption that Bob knows the contract before announcing any forecasts can be replaced with the assumption that Bob correctly anticipates the future rewards and penalties. However if Bob is uninformed, and also ignorant regarding the future rewards and penalties, then there exist verification contracts which screen informed and uninformed experts.

C. *Refutation Contracts*

We now return to the case of unrestricted theories, where verification contracts do not deliver an effective way to determine whether Bob's theory is based on any relevant knowledge of the data-generating process. However, in this section, we consider refutation contracts.

Proposition 4. *Assume that theories are not restricted. Then there exists a refutation contract (with bounded payoffs) that screens between informed and uninformed experts.*

Propositions 2, 3, and 4 show a basic asymmetry between verification and refutation contracts: no verification contract can screen between informed and uninformed experts, but some refutation contracts can. In order to prove proposition 4, we now construct a screening refutation contract. This contract is constructed in the context of Popper's main example.

Assume that in every period a swan is observed. This swan can be white or of another color. Let 1 denote white color, and 0 any other color. Let 1_t be the $(t+1)-$history of white swans in all periods until period t, but a non-white swan at period $t+1$. Let $\bar{1}_m$ consist of the union of 1_t, $t \geq m$. So, $\bar{1}_m$ are the histories in which only white swans are seen for at least m periods (starting at the initial period) after which a nonwhite swan is observed in the period following the initial sequence of white swans.

Now, we return to the classic induction problem: a long sequence of white swans does not prove that all swans are white. For some data-generating processes, a long consecutive sequence of white swans may be followed by a nonwhite with positive probability. However, for every process, the pattern of a sufficiently long consecutive sequence of white swans followed by a nonwhite swan is unlikely. This can be shown as follows: note that $\bar{1}_m \downarrow \varnothing$ as m goes to infinity (because $\bar{1}_{m+1}$ is contained in $\bar{1}_m$ and the interception of all sets $\bar{1}_m$ is empty). So, for every probability measure P, $P(\bar{1}_m) \downarrow 0$ as m goes to infinity. Hence, for *any* data-generating process, the pattern $\bar{1}_m$ (in which only white swans are seen until period $t \geq m$ and a nonwhite swan is seen at period $t+1$) is unlikely if m is large enough.

We define a contract in which Bob pays if he announces a theory f and Alice observes the pattern of more than $m(f)$ consecutive white swans followed by a nonwhite swan, where $m(f)$ is long enough so that this pattern is unlikely, i.e., $P_f(\bar{1}_{m(f)}) \leq \varepsilon$, $\varepsilon > 0$. So, the Cournot principle is implicitly used here because negative payoffs are triggered not by conclusive rejection, but by the observation of events deemed unlikely by the theory.

Consider the following contract: At period 0, Bob receives $\delta \in (\varepsilon, 0.5)$ for announcing any theory f. Bob also receives disutility 1 contingent on $\bar{1}_{m(f)}$, and no disutility otherwise. Formally,

$$\begin{aligned} u(f, \varnothing) &= \delta; \\ u(f, s_t) &= -1 \quad \text{if} \quad s_t \in \bar{1}_{m(f)}; \\ u(f, s_t) &= 0 \quad \text{if} \quad s_t \notin \bar{1}_{m(f)}. \end{aligned} \tag{21}$$

Call this contract a *simple refutation contract.*

This contract is an example of a *test-based contract*, in which a set of finite histories R_f (called the rejection set) is defined as inconsistent with theory f, and Bob incurs a disutility of 1 if the observed history belongs to R_f, i.e., is inconsistent with his theory. The parallel to testing is clear: 1 is a disutility incurred when the theory is rejected. In the simple refutation contract, R_f is defined as $\bar{1}_{m(f)}$.

The properties of this contract can be checked as follows: first note that

$$U(f,s) = \begin{cases} \delta - 1 & \text{if for some period } t,\ s_t \in \bar{1}_{m(f)},\ s = (s_t, ...); \\ \delta & \text{if for every period } t,\ s_t \notin \bar{1}_{m(f)},\ s = (s_t, ...). \end{cases} \tag{22}$$

Bob's utility function U takes only two values. Bob can be punished (i.e., receive a negative payoff) only once. Punishment occurs if after period $m(f)$, a nonwhite swam is observed for the first time. Nonwhite swans observed for the second time (or before $m(f)$) do not trigger a negative payoff. Hence, the assumption of bounded utility functions u and U is satisfied.

Assume that Bob claims (based on a theory $\bar{f}$) that all swans are white. The simple refutation contract then punishes Bob only if a nonwhite swan is observed. This intuitive property suffices for payoffs $U(f,s)$ *not* to be lower semi-continuous in s. To see this, consider the path $s = (1, 1, ...)$ in which only white swans are observed. Then, $U(\bar{f}, s) = \delta$ because Bob is never punished. Now consider any path $s(t)$ such that the first t outcomes are white swans and the $(t+1)$-st outcome is a nonwhite swan. By definition, $s(t)$ converges to s as t goes to infinity, but $U(\bar{f}, s(t)) = \delta - 1$ is significantly lower than $U(\bar{f}, s) = \delta$. As we argued in section III A, lower semi-continuity is the critical feature of verification experts that makes them incapable of screening between informed and uninformed contracts. However, lower semi-continuity is also a strong condition that is naturally denied in several refutation contracts. Informally, lack of semi-continuity may hold because a single observation may not change a large data set topologically, but may have a strong negative impact on the evaluation of some theories.

An informed expert accepts the simple refutation contract because if the data-generating process is announced, the punishment is unlikely. More precisely,

$$\bar{U}^{P_f}(f) = \delta - P_f(\bar{1}_{m(f)}) \geq \delta - \varepsilon > 0. \tag{23}$$

On the other hand, an uninformed expert turns down the simple refutation contract. Let $\mathcal{T}_m \subseteq \mathcal{T}$ be the set of all theories such that $\bar{1}_m \subset R_f$, i.e.,

$$\mathcal{T}_m = \{ f \in \mathcal{T} : \bar{1}_m \subset R_f \}. \tag{24}$$

That is, $\mathcal{T}_m$ are the theories that may be rejected at period m. Since $R_f = \bar{1}_{m(f)}$, and $\bar{1}_{m+1} \subseteq \bar{1}_m$, it follows that $\mathcal{T}_m \subseteq \mathcal{T}_{m+1}$. Moreover, any theory $f \in \mathcal{T}$ belongs to $\mathcal{T}_m$ for some m, because $m \geq m(f)$ for sufficiently large values of m; as a result, $\mathcal{T}_m \uparrow \mathcal{T}$ as m goes to infinity. That is, as the period m increases, more theories can be rejected, and every theory can be rejected at some point in time. Thus, for every random generator of theories $\zeta \in \Delta(\mathcal{T})$, there exists a period m^* such that

$$\zeta(\mathcal{T}_{m^*}) \geq 1 - \delta. \tag{25}$$

Indeed, $\mathcal{T}_m \uparrow \mathcal{T}$ implies $\zeta(\mathcal{T}_m) \uparrow \zeta(\mathcal{T}) = 1$. If $s = (1_{m^*}, ...)$, then

$$E^{\zeta}U(f, s) = \delta - \zeta(\mathcal{T}_{m^*}) \leq -1 + 2\delta < 0, \tag{26}$$

since $\delta < 0.3$. Hence, $V(\zeta) < 0$ for all random generators of theories $\zeta \in \Delta(\mathcal{T})$. This shows that an uninformed expert turns down this contract, completing the proof of proposition 4.

The simple refutation contract screens between informed and uninformed experts. In conjunction with inability of verification contracts to do the same, this delivers an original argument in support of the idea that refutation is a better guiding principle for empirical research than verification. At the core of this original argument is this fundamental contrast between refutation contracts and verification contracts: there exists a refutation contract that can screen between informed and uninformed experts, but no verification contract can screen between these experts.

D. *Refutation and Exchangeability*

Propositions 2, 3, and 4 show an asymmetry between verification and refutation when theories are not restricted. Now consider the case such that theories must be exchangeable.

There is an important difference between the exchangeable and the unrestricted cases: when processes are restricted to be exchangeable, it may be possible to learn future odds from the observed data, but learning may not be possible when processes are unrestricted. So, in the exchangeable case, both the expert and the tester will eventually know the approximate odds of future events. However, if no restrictions are placed a priori on the data-generating process, then there is no (known) way of making accurate statistical inferences from the data. Despite this fundamental difference, a screening refutation contract exists both in the case of unrestricted theories and also in case of the restricted theories to being exchangeable.

Proposition 5. *The simple refutation contract, with theories restricted to be exchangeable, still screens informed and uninformed experts.*

Propositions $2 - 5$ deliver a conceptual differentiation between verification and refutation. Both in the case of unrestricted theories and also in the case of exchangeable theories, no verification contract can screen between informed and uninformed experts, but some refutation contracts can.

The proof of proposition 5 is as follows: Consider the simple refutation contract, with theories restricted to be exchangeable. An informed expert accepts it because, by proposition 4, any informed expert accepts it. Now consider an uninformed expert who produces theories with some random generator of theories ζ. As in the proof of proposition 4, let m^* be given by (25). Consider now the *i.i.d.* process $P^P \in \Delta(\Omega)$ (hence, an exchangeable process) such that the probability p of a white swan satisfies

$$p^{m^*} > \frac{\delta}{1-\delta}. \tag{27}$$

Then,

$$P^p(\bar{1}_{m^*}) = \sum_{n=m^*}^{\infty} p^n(1-p) = p^{m^*} > \frac{\delta}{1-\delta}. \tag{28}$$

Thus,

$$E^{P^P} E^{\zeta} U(f,s) \leq (1 - P^p(\bar{1}_{m^*}))\delta + P^p(\bar{1}_{m^*})(-1+\delta) < 0. \tag{29}$$

So, for some exchangeable process, the uninformed expert total payoff is negative. Hence, the uninformed expert turns down this contract.

It follows from the proof of propositions 4 and 5 that the simple refutation contract screens informed and uninformed experts when theories are restricted to any class which contains all deterministic processes (e.g., falsifiable theories) and also to any class of processes that contains all *i.i.d.* processes.[15] Hence, the asymmetry between verification and refutation extends to several other classes of theories as well (see the De Finetti's type results in Diaconis and Freedman (1987)). However, as mentioned in section III B, for some restrictions on the set of allowed theories, no screening contracts exists. For other restrictions, there are both refutation and verification screening contracts.

E. *Additional Properties, and Drawbacks of the Simple Refutation Contract*

Olszewski and Sandroni (2008), and Eran Shmaya (2008) consider a large class of empirical tests ordinarily used in statistics (such as calibration and likelihood tests). Like verification contracts, the contracts based on these tests cannot screen between informed and uninformed experts even if theories are restricted to convex and closed sets. Therefore the simple refutation contract and the contract in proposition 5 are original ways of testing theories. That is, the criteria we propose do not follow from standard results in statistics. Moreover, since ε (and hence, δ) can be made arbitrarily small, it follows that Alice need only make a small payments to induce an informed expert to accept the simple contract. Finally, if ε is very small, then it is almost optimal for Bob, if informed, to reveal his theory truthfully. This follows because the odds that Bob will incur any disutility can be made arbitrarily small (as long as he is informed and truthfully reveals his theory).

[15] It is an open question whether propositions 2 and 3 still hold if theories are restricted to be falsifiable.

Finally, we wish to point out that although our results deliver a basic conceptual distinction between verification and refutation, but we do not claim that our refutation contracts are immune to either practical or conceptual shortcomings. One weakness of our tests, and of the contracts based on them, is that an uninformed expert fails the test only in the case of few histories or processes. However, in the case of unrestricted theories only, Olszewski and Sandroni (2009b) show a test-based refutation contract (that can screen between informed and uninformed experts) such that an uninformed expert fails the test, no matter how he randomizes, on a topologically large set of histories. (See Dekel and Feinberg (2006) for an earlier work on screening tests, and a follow-up work by Shmaya (2008) for other test-based refutation contracts that can screen between informed and uninformed experts; all these contracts are far more complex than the simple refutation contract. See also Al-Najjar, Smorodinsky, Sandroni, and Weinstein (2009) for a very recent work on restricted theories to a particular, non-closed set.) There are also additional difficulties with refutation tests, such as the need for long data sets to reject theories, but these are beyond the scope of this paper.

F. *Contracts with arbitrary transfers*

One could conjecture that verification and refutation contracts are equivalent, because one may, for example, raise utilities $u(f, s_t)$ by some constant and then compensate for this change with a change in the initial utility $u(f, \emptyset)$. There are, however, some difficulties with this idea. Consider our simple refutation contract and assume that we raise all utilities $u(f, s_t)$ by 1. Now Bob receives positive payoff in every period such that some number of white swans followed by a nonwhite swan is not observed, and no payoff if such a pattern is observed. Then, along the path $s = (1, 1, ...)$, where only white swans are seen, Bob's contingent net payoff at period zero becomes infinite.

However, every screening contract can be transformed into a refutation screening contract. Let $U_{C_i}(f, s)$ denote Bob's contingent payoff at period zero under contract C_i, where $i = 1, 2$. We say that contract C_2 is *less beneficial* (to Bob) than contract C_1 if

$$U_{C_2}(f, s) \leq U_{C_1}(f, s) \tag{30}$$

for all data sequences s and theories f.

Proposition 6. *For every contract C_1 which is accepted by an informed expert, there exists a less beneficial refutation contract C_2 that is also accepted by an informed expert.*

So, if an uninformed expert rejects contract C_1, then he will also reject contract C_2. This result holds no matter how theories are restricted. In particular, no matter what the restrictions on theories, if a screening verification contract exists, then so does a screening refutation contract.

IV. Conclusion

Falsifiability is a widely used guide in research and legal proceedings because it is perceived as a requirement that could disqualify nonscientific theories. Indeed, falsifiable theories can be conclusively rejected, whereas nonfalsifiable ones cannot. However, we show that falsifiability imposes essentially no constraints when theories are produced by strategic experts. Without any knowledge, it is possible to construct falsifiable theories that are unlikely to be falsified, no matter how the data unfold in the future.

Verification suffers from the same difficulty as falsification. Strategic experts, with no knowledge of the data-generating process, can produce theories that are likely to turn out consistent with the data. However, there are special ways of constructing refutation contracts (by defining which data are inconsistent with each theory) that can screen legitimate from worthless theories, even if experts are strategic.

APPENDIX

A. *Proof of Proposition 1*

A cylinder with base on history $s_t = (s^1, ..., s^t)$ is denoted by $C(s^1, ..., s^t)$. Take any $\varepsilon > 0$. We will construct a $\bar{\zeta} \in \Delta(\mathcal{F})$ such that for every path $(s^1, s^2, ...) \in \Omega$,

$$\bar{\zeta}\left\{f \in \mathcal{F} \mid \exists_t \quad (s^1, ..., s^t) \in R_f \right\} < \varepsilon. \tag{A1}$$

This will complete the proof since

$$V(\bar{\zeta}) \geq (1-\varepsilon)u + \varepsilon(u-d) > u - \rho \tag{A2}$$

if ε is sufficiently small.

Take a number $r > 0$ so small that

$$\sum_{t=1}^{\infty} r^t < \varepsilon, \tag{A3}$$

and next take a sequence of natural numbers $\{M_t,\ t = 1, 2...\}$ such that

$$\frac{1}{n^{M_t+1}} < r^t. \tag{A4}$$

It will be convenient to represent Ω as the Cartesian product $\prod_{t=1}^{\infty} X_t$ of sets $X_t = S^{M_t}$, $t = 1, 2, ...$ Consider a sequence of independent random variables $\tilde{X}_t$ uniformly distributed on the set X_t. Let $\tilde{X}$ be the product $\prod_{t=1}^{\infty} \tilde{X}_t$ of random variables $\tilde{X}_t$, $t = 1, 2, ...$; that is, $\tilde{X} = (x_1, ..., x_t, ...)$, $x_t \in X_t$, if and only if $\tilde{X}_t = x_t$ for all $t = 1, 2, ...$

Let

$$Z_t = \sum_{j=1}^{t} M_j. \tag{A5}$$

For $x = (x_1, ..., x_t, ...) \in \Omega$, let

$$\overline{S}_x = C(x_1) \cup \bigcup_{t=1}^{\infty} \bigcup_{z_t \in S^{Z_t}} C(z_t, x_{t+1}) \tag{A6}$$

be the union of the cylinders with base on histories of the form (z_t, x_{t+1}), where z_t is an arbitrary element of S^{Z_t} and x_{t+1} are $Z_t + 1, ..., Z_{t+1}-$period outcomes of infinite history x. Let f_x be the theory which assigns equal probabilities to all outcomes $s \in S$, contingent on all finite histories, except histories $(z_t, x_{t+1}) \in \overline{S}_x$. Contingent on histories $(z_t, x_{t+1}) \in \overline{S}_x$, let theory f_x assign equal probabilities to all outcomes $s \in S$ except one, denote by ω_{t+1}, to which it assigns probability 0.

Observe that all theories f_x, $x \in \Omega$, are falsifiable. Indeed, for any history $(s^1, ..., s^t) \in S^t$, take a $Z_m \geq t$ and any extension $z_m \in S^{Z_m}$ of $(s^1, ..., s^m)$. Then (z_m, x_{m+1}) is also an extension of $(s^1, ..., s^t)$, and by definition, $(z_m, x_{m+1}) \in \overline{S}_x$. Thus, f_x assigns probability 0 to outcome ω_{m+1} contingent on history (z_m, x_{m+1}). So, $P_{f_x}(z_m, x_{m+1}, \omega_{m+1}) = 0$.

Let $\bar{\zeta} \in \Delta(\mathcal{F})$ be defined as follows. First a realization x of the random variable $\tilde{X}$ is observed, and then theory f_x that is announced.

Fix a path $y = (s^1, s^2, ...) \in \Omega$; it can also be represented as $y = (y_1, ..., y_t, ...)$ where $y_t \in X_t$; so, y_1 consists of the first M_1 outcomes of $(s^1, s^2, ...)$, and y_{t+1} consists of $Z_t + 1, ..., Z_{t+1}-$period outcomes. By definition, $(y_1, ..., y_t, ...) \in R_{f_x}$ if and only if $y_{t+1} = x_{t+1}$ for some $t \in 0, 1, ...$, and $s^m = \omega_{t+1}$ in period $m = Z_{t+1} + 1$. Hence,

$$\bar{\zeta}\left\{f \in \mathcal{F} \mid \exists_t \quad (s^1, ..., s^t) \in R_f \right\} = \frac{1}{n}\hat{P}\left(\overline{S}_y\right), \tag{A7}$$

where $\hat{P}$ is the uniform probability distribution on Ω.

Since

$$\hat{P}\left(\overline{S}_y\right) \leq \sum_{t=0}^{\infty} \hat{P}\left(\bigcup_{z_t \in S^{Z_t}} C(z_t, x_{t+1})\right) = \sum_{t=0}^{\infty} \frac{1}{n^{M_{t+1}}} \leq n \sum_{t=1}^{\infty} r^t, \tag{A8}$$

$$\bar{\zeta}\left\{f \in \mathcal{F} \mid \exists_t \quad (s^1, ..., s^t) \in R_f \right\} \leq \sum_{t=1}^{\infty} r^t < \varepsilon.$$

B. *Proof of Propositions 2 and 3*

Let X be a metric space. Recall that a function $g : X \to R$ is *lower semi-continuous* at an $x \in X$ if for every sequence $(x_n)_{n=1}^{\infty}$ converging to x:

$$\forall_{\varepsilon>0} \quad \exists_{\overline{n}} \quad \forall_{n \geq \overline{n}} \quad g(x_n) > g(x) - \varepsilon. \tag{A9}$$

The function g is lower semi-continuous if it is lower semi-continuous at every $x \in X$.

As mentioned in footnote 6, we endow $\Delta(\Omega)$ with the weak$-*$ topology and with the $\sigma-$algebra of Borel sets, (i.e., the smallest $\sigma-$algebra which contains all open sets in weak$-*$ topology). We endow the set of all theories $\mathcal{T}$ with the pointwise convergence topology; in this topology, a sequence of theories $(f_n)_{n=1}^{\infty}$ converges to a theory f if $f_n(s_t) \to_n f(s_t)$ for every history $s_t \in \overline{S}$. Let $\Delta(\mathcal{T})$ be the set of probability measures on $\mathcal{T}$. We also endow $\Delta(\mathcal{T})$ with the weak$-*$ topology. It is well-known that Ω, $\Delta(\Omega)$, $\mathcal{T}$, and $\Delta(\mathcal{T})$ are compact metrizable spaces.

Recall that if $\Delta(X)$ is endowed with the the weak$-*$ topology, then if $h : X \to R$ is a continuous function, then $H : \Delta(X) \to R$ defined by $H(\mu) = E^{\mu}(h)$, where E^{μ} is the expectation operator associated with $\mu \in \Delta(X)$, is also a continuous function.

Lemma A1. Assume that an informed expert accepts a verification contract. Then, there exists $\delta > 0$ such that for every probability measure $P \in \Delta(\Omega)$ there exists a theory $f(P)$ such that $\bar{U}^P(f(P)) \geq \delta$.

Proof : Suppose, by contradiction, that there exists a sequence of probability measures P_1, P_2,....such that $\bar{U}^{P_i}(f) \leq \frac{1}{i}$ for every theory f. Let f_i be a theory such that $P_{f_i} = P_i$, $i = 1, 2, ...$ Let $\bar{f}$ and $P_{\bar{f}}$ be the limits of some subsequences (also indexed by i). So, $f_i \to_i \bar{f}$ and $P_i \to_i P_{\bar{f}}$ as i goes to infinity. It follows that $\bar{U}^{P_i}(\bar{f}) \leq 1/i$. We now show that $\bar{U}^{P_{\bar{f}}}(\bar{f}) \leq 0$.

Let $X_t(\bar{f}, s) = u(\bar{f}, s_t)$, where $s = (s^1, ..., s^t, ...)$, for $t \geq 1$, and let $X_0(\bar{f}, s) = u(\bar{f}, \varnothing)$. Then,

$$U(\bar{f}, s) = \sum_{t=0}^{\infty} X_t(\bar{f}, s). \tag{A10}$$

By the monotone convergence theorem,

$$E^{P_{\bar{f}}} \left\{ \sum_{t=0}^{m} X_t(\bar{f}, s) \right\} \underset{m \to \infty}{\longrightarrow} \bar{U}^{P_{\bar{f}}}(\bar{f}). \tag{A11}$$

Since each $X_t(\bar{f}, s)$ is a continuous function of s, by the definition of the weak$-*$ topology,

$$E^{P_i} \left\{ \sum_{t=0}^{m} X_t(\bar{f}, s) \right\} \underset{n \to \infty}{\longrightarrow} E^{P_{\bar{f}}} \sum_{t=0}^{m} X_t(\bar{f}, s) \tag{A12}$$

for every m.[16] In addition,

$$E^{P_i} \left\{ \sum_{t=0}^{m} X_t(\bar{f}, s) \right\} \leq \bar{U}^{P_i}(\bar{f}) \leq \frac{1}{i}. \tag{A13}$$

Hence,

$$E^{P_{\bar{f}}} \left\{ \sum_{t=0}^{m} X_t(\bar{f}, s) \right\} \leq 0 \tag{A14}$$

[16] Function $X_t(\bar{f}, s)$ is continuous in s, because its value depends only on s_t, and any function with this property is continuous.

for every m. It now follows that $\bar{U}^{P_{\bar{f}}}(\bar{f}) \leq 0$, this is a contradiction with the assumption $\bar{U}^{P_f}(f) > 0$ for every theory f.

Consider an arbitrary verification contract. Let $\mathcal{H} : \Delta(\mathcal{T}) \times \Delta(\Omega) \longrightarrow \Re$ be a function defined by

$$\mathcal{H}(\zeta, P) = E^{\zeta} E^{P} U(f, s). \tag{A15}$$

Step 1: Assume that $\mathcal{H}(\zeta, P) < \infty$ for every $\zeta \in \Delta(\mathcal{T})$ and $P \in \Delta(\Omega)$, and that for every $s_t \in \bar{S}$, $u(f, s_t)$ is a bounded function of f. Then, for every $\zeta \in \Delta(\mathcal{T})$, $\mathcal{H}(\zeta, P)$ is a lower semi-continuous function of P.

Proof: Let again $X_t(f, s) = u(f, s_t)$ where $s = (s^1, ..., s^t, ...)$, $t \geq 1$, and $X_0(f, s) = u(f, \varnothing)$. Then,

$$U(f, s) = \sum_{t=0}^{\infty} X_t(f, s). \tag{A16}$$

By the monotone convergence theorem,

$$E^P U(f, s) = \sum_{t=0}^{\infty} E^P X_t(f, s) \text{ and } \mathcal{H}(\zeta, P) = \sum_{t=0}^{\infty} E^{\zeta} E^P X_t(f, s). \tag{A17}$$

Fix $\varepsilon > 0$ and assume that P_n converges to P (in the weak$-^*$ topology) as n goes to infinity. Then, by the definition of the weak$-^*$ topology, $E^{P_n} X_t(f, s) \underset{n \to \infty}{\longrightarrow} E^P X_t(f, s)$ for every t. By the dominated convergence theorem, and the assumption that $u(f, s_t)$ is a bounded function of f, $E^{\zeta} E^{P_n} X_t(f, s) \underset{n \to \infty}{\longrightarrow} E^{\zeta} E^P X_t(f, s)$ for every t.

Now, since $\mathcal{H}(\zeta, P) < \infty$, there exists m^* such that

$$\sum_{t=m^*+1}^{\infty} E^{\zeta} E^P X_t(f, s) < \frac{\varepsilon}{2}. \tag{A18}$$

If n is sufficiently large,

$$\sum_{t=0}^{m^*} E^{\zeta} E^{P_n} X_t(f, s) \geq \sum_{t=0}^{m^*} E^{\zeta} E^P X_t(f, s) - \frac{\varepsilon}{2} \tag{A19}$$

and, therefore,

$$\mathcal{H}(\zeta, P_n) \geq \sum_{t=0}^{m^*} E^{\zeta} E^{P_n} X_t(f, s) \geq \mathcal{H}(\zeta, P) - \varepsilon. \tag{A20}$$

Step 2: The sets $\Delta(\mathcal{T})$ and $\Delta(\Omega)$ are compact sets in the weak$-^*$ topology and $\mathcal{H}(\zeta, P)$ is a linear function on ζ and P. Hence, it follows from Fan's (1953) minmax theorem that if $\mathcal{H}(\zeta, P)$ is lower semi-continuous in P, then

$$\inf_P \sup_{\zeta} \mathcal{H}(\zeta, P) = \sup_{\zeta} \inf_P \mathcal{H}(\zeta, P). \tag{A21}$$

Proof of proposition 2: For a given P take an $f(P)$ such that $\bar{U}^P(f(P)) \geq \delta$. The existence of an $f(P)$ with this property is guaranteed by Lemma $A.1$. Take ζ such that $\zeta(\{f(P)\}) = 1$. Then, the left-hand side of (A21) exceeds δ. So, the right-hand side of (A21) exceeds δ. This yields $\bar{\zeta} \in \Delta(\Delta(\Omega))$ such that $\mathcal{H}(\bar{\zeta}, P) > \delta/2$ for every P. Now, for any $s \in \Omega$, take the probability measure P_s such that $P_s(\{s\}) = 1$. So, $\mathcal{H}(\bar{\zeta}, P_s) > \delta/2$ for every s. Hence, $V(\bar{\zeta}) > \delta/2 > 0$.

Proof of proposition 3: For a given P take a ζ such that $\zeta(\{f\}) = 1$ for some f such that $P_f = P$. Then, the left-hand side of (A21) exceeds $V(I, c) + c$. So, the right-hand side of (A21) exceeds $V(I, c) + c$. This yields $\bar{\zeta} \in \Delta(\Delta(\Omega))$ such that $\mathcal{H}(\bar{\zeta}, P_s) > V(I, c) + 0.5c$ for every $s \in \Omega$. Hence, $V(\bar{\zeta}) \geq V(I, c) + 0.5c > V(I, c)$.

Remark 1. *Proposition 2 and 3 are shown in the case of unrestricted theories. The proof for the case of Θ closed, compact is, however, very similar. The critical property of $\Delta(\Omega)$ used in the proofs is that $\Delta(\Omega)$ is compact and convex set.*

C. *Proof of Proposition 6*

Given a sequence of numbers a_n and a sequence of pairwise disjoint measurable sets $A_n \subset \Omega$, where $n = 1, ..., N$, define a step-wise function $Y : \Omega \to R$ by

$$Y(s) = \sum_{n=1}^{N} a_n I_{A_n}, \tag{A22}$$

where I_{A_n} is the indicator function of A_n, i.e.,

$$I_{A_n}(s) = \begin{cases} 1 \text{ if } s \in A_n \\ 0 \text{ if } s \notin A_n \end{cases}. \tag{A23}$$

Fix a theory f, let $P = P_f$, and let $X(s) = U_{C_1}(s)$. Since X is a measurable and bounded function of s, and $E^P X(s) > 0$, there exist numbers a_n and pairwise disjoint measurable sets A_n such that

$$Y(s) \leq X(s) \text{ for every } s \in \Omega, \text{ and } E^P Y(s) > 0. \tag{A24}$$

Let $b = \max\{a_1, ..., a_N\}$, and let $b_n = b - a_n$, $n = 1, ..., N$. Then

$$Y(s) = b - \sum_{n=1}^{N} b_n I_{A_n}. \tag{A25}$$

Since sets A_n are measurable, for any $\rho > 0$, there exist an open sets $U_n \subset \Omega$ such that $A_n \subset U_n$ and $P(U_n) < P(A_n) + \rho$.[17] This yields that there exist open sets $U_n \subset \Omega$

[17] This is the place in the proof in which it is essential that we construct a *refutation* contract. If we were trying to prove, in a similar manner, the counterpart of proposition 6 for *verification* contracts, we would have to find open sets $U_n \subset \Omega$ such that $U_n \subset A_n$ and $P(A_n) < P(U_n) + \rho$, but such sets may not exist in general.

such that $A_n \subset U_n$ and $E^P Z(s) > 0$, where

$$Z(s) = b - \sum_{n=1}^{N} b_n I_{U_n}; \tag{A26}$$

of course, since $b_n \geq 0$, $n = 1, ..., N$, $Z(s) \leq Y(s) \leq X(s)$ for every $s \in \Omega$.

Each open set U_n can be represented as a union of pairwise disjoint cylinders C_n^k. Define contract C_2 as follows: let

$$u_{C_2}(f, \varnothing) = b, \tag{A27}$$

and for every finite history $\overline{s} \in \overline{S}$, let

$$u_{C_2}(f, \overline{s}) = \sum_{n=1}^{N} \left\{ -b_n : C_n^k = C(\overline{s}) \text{ for some } k \right\}. \tag{A28}$$

That is, Bob receives payment b when he announces a theory, and then he has to pay b_n when history $\overline{s}$ that is the base of a cylinder C_n^k is observed.

By definition, C_2 is a refutation contract, and

$$U_{C_2}(f, s) = Z(s) \text{ for every } s \in \Omega. \tag{A29}$$

Thus, contract C_2 is less beneficial than contract C_1, but is still accepted by an informed expert.

REFERENCES

Ahn, David S. 2008. "Ambiguity without a State Space." *Review of Economic Studies*, 75(1): 3-28.

Al-Najjar, Nabil I. and Jonathan L. Weinstein. 2008. "Comparative Testing of Experts." *Econometrica*, 76(3): 541-559.

Al-Najjar, Nabil I., Rann Smorodinsky, Alvaro Sandroni, and Jonathan L. Weinstein. 2009. "Testing Theories with Learnable and Predictive Representations." Unpublished.

Carvajal, Andrés, Indrajit Ray, and Susan K. Snyder. 2004. "Equilibrium Behavior in Markets and Games: Testable Restrictions and Identification." *Journal of Mathematical Economics*, 40(1-2): 1-40.

Cesa-Bianchi Nicolò and Gábor Lugosi. 2006. *Prediction, Learning and Games*. Cambridge: Cambridge University Press.

Cournot, Antoine A. 1843. *Exposition de la Théorie des Chances et des Probabilités*. Paris: Librairie de *L.* Hachette.

De Finetti, Bruno. 1937. "La Prévision: ses lois Logiques, ses Sources Subjectives." *Annales de l'Institut Henri Poincaré*. 7(1): 1-66.

Dekel, Eddie and Yossi Feinberg. 2006. "Non-Bayesian Testing of a Stochastic Prediction." *Review of Economic Studies*, 73(4): 893 - 906.

Diaconis, Persi, and David Freedman. 1987. "A Dozen De Finetti-style Results in Search of a Theory." *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*: 23: 397-423.

Fan, Ky. 1953. "Minimax Theorems." *Proceedings of the National Academy of Sciences U.S.A.,* 39(1): 42-47.

Feinberg, Yossi, and Colin Stewart. 2008. "Testing Multiple Forecasters." *Econometrica*, 76(3): 541-582.

Fortnow, Lance J., 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): 379-390.

Fudenberg, Drew, and David K. Levine. 1999. "An Easier Way to Calibrate." *Games and Economic Behavior*, 29(1-2): 131-137.

Gajdos, Thibault, Takashi Hayashi, Jean-Marc Tallon, and Jean-Christophe Vergnaud. 2008. "Attitude toward Imprecise Information." *Journal of Economic Theory*, 140(1): 27-65.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maxmin Expected Utility with A Non-Unique Prior." *Journal of Mathematical Economics*, 18(2): 141–153.

Hart, Sergiu, and Andreu Mas-Colell. 2001. "A General Class of Adaptative Strategies." *Journal of Economic Theory*, 98(1): 26-54.

Kreps, David M. 1988. *Notes on the Theory of Choice*. Boulder: Westview Press.

Kalai, Ehud, Ehud Lehrer, and Rann Smorodinsky. (1999) "Calibrated Forecasting and Merging." *Games and Economic Behavior*, 29(1-2): 151-169.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. Boston: Houghton Mifflin Co.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.

Lakatos, Imre, and Alan Musgrave, ed. 1970. *Criticism and the Growth of Knowledge*. Vol.4. Cambridge: Cambridge University Press.

Lehrer, Ehud. 2001. "Any Inspection Rule is Manipulable," *Econometrica*, 69(5): 1333-1347.

Lehrer, Ehud, and Eilon Solan. 2003. "No Regret with Bounded Computational Capacity." Unpublished.

LeRoy Stephen F., and Larry D. Singell, Jr. 1987. "Knight on Risk and Uncertainty." *Journal of Political Economy*, 95(2): 394-406.

Olszewski, Wojciech. 2007. "Preferences over Sets of Lotteries." *Review of Economic Studies,* 74(2): 567-595.

Olszewski, Wojciech, and Alvaro Sandroni. 2007. "Contracts and Uncertainty." *Theoretical Economics*, 2: 1-13.

Olszewski, Wojciech, and Alvaro Sandroni. 2008. "Manipulability of Future-Independent Tests" *Econometrica*, 76(6): 1437-1466.

Olszewski, Wojciech, and Alvaro Sandroni. 2009a. "Manipulability of Comparative Tests." *Proceedings of the National Academy of Sciences U.S.A.*, 106(13): 5029-5034.

Olszewski, Wojciech, and Alvaro Sandroni. 2009b. "A Nonmanipulable Test." *Annals of Statistics,* 37(2): 1013 -1039.

Popper, Karl R. 1968. *The Logic of Scientific Discovery*. London: Hutchinson & Co.

Reny, Philip J. 1999. "On the Existence of Pure and Mixed Strategy Nash Equilibria in Discontinuous Games." *Econometrica*, 67(5): 1029-1056

Rustichini, Aldo. 1999. "Minimizing Regret: The General Case." *Games and Economic Behavior*, 29(1-2): 224-243.

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

Sandroni, Alvaro, Rann Smorodinsky and Rakesh V. Vohra. 2003. "Calibration with Many Checking Rules." *Mathematics of Operations Research*, 28(1): 141-153.

Shmaya, Eran. 2008. "Many inspections are manipulable." *Theoretical Economics*, 3(3): 367–382.

Vovk, Vladimir, and Glenn Shafer. 2005. "Good Randomized Sequential Probability Forecasting is Always Possible," *Journal of the Royal Statistical Society Series B*, 67(5): 747 - 763.