Mathematical Social Sciences 65 (2013) 5–9

Contents lists available at SciVerse ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

A difficulty in the testing of strategic experts

Nabil Al-Najjar, Alvaro Sandroni *

Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston IL 60208, United States

ARTICLE INFO

Article history:
Received 8 November 2011
Received in revised form
30 April 2012
Accepted 24 July 2012
Available online 3 August 2012

ABSTRACT

We consider a contracting problem between a principal who wants to be informed about relevant stochastic processes and an expert who claims to know which process will generate the data. The data generating process is known to belong to a given class.

We show that if the expert discounts the future and the set of allowed processes is convex then there is no screening contract that separates informed and uninformed experts. Our main proviso of convexity is immediately satisfied by any class of processes that can be characterized in a De Finetti-style result. This proviso is also satisfied when the expert is required to produce a prior over the relevant parameter space. Thus, the main difficulty in screening informed and uninformed experts has not yet been fully resolved.

1. Introduction

Testing theories is an essential activity in science. The standard methodology for testing theories in economics (and in science in general) assumes, under a null hypothesis, that the theory is correct. The theory is rejected when data deemed to be inconsistent with the null hypothesis is observed.

These standard statistical methods may be satisfactory when experts are assumed to be honest. However, economists and practitioners alike have long expressed doubts on the assumption that agents will report what they know honestly unless they are given the proper incentives to do so. Experts may have reputational concerns and may misrepresent what they know in order to maintain a (perhaps false) reputation of knowledge. So, a natural question in economics is how to develop empirical tests that allow for the possibility of strategic experts with reputational concerns.

A growing literature in economics uses the following set-up to allow for strategic experts: A principal, named Alice, hires a potential expert, named Bob, who claims to be informed about a relevant stochastic process. Alice faces an adverse selection problem: Bob knows his type (i.e., whether he is informed about the relevant process), but Alice does not know whether Bob does, in fact, know anything significant about the probabilities of interest (i.e., something she does not know). A natural conjecture is that Alice can mitigate her adverse selection problem by showing Bob a contract which specifies payments depending on the process Bob announced at period zero and the future realization of the data. Empirical tests are a special case of contracts where Bob receives a reward when he announces his process and a penalty if the data deemed inconsistent with the announced process is observed (or alternatively a reward if the data deemed consistent with he announced process is observed).

A screening contract is such that Bob, if informed, accepts and Bob, if uninformed, rejects. A special case is a test that has two basic properties: (I) Bob is very likely to pass the test if informed and (II) If uninformed, Bob is uncertain whether he can produce processes strategically to pass the test. Property I is the fundamental requirement of statistical methods mentioned in our opening paragraph. Property II must hold otherwise the uninformed expert will not fear rejection and accept the test.

A screening contract may inform Alice whether or not Bob is informed. So, a basic question is whether there exists a screening contract. A series of papers answer this question in the negative (see our literature review). In particular, Olszewski and Sandroni (2007) show that if a perfectly informed expert (i.e., announces the actual data generating process) accepts Alice's contract then a completely uninformed expert (i.e., knows nothing about the data generating process) also accepts it. This holds even if Alice is willing to pay large amounts to learn the relevant probabilities and the uninformed expert is so averse to uncertainty that he only accepts the contract when he is better off with it, no matter which permissible process generates the data.

The impossibility results often assume no restrictions over the theories the expert may announce. So, it is tempting to speculate that this absence of constraints plays a central role in these results. The absence of constraints may give Bob sufficient flexibility to produce a process that will eventually yield sufficiently high payoffs to him, no matter how the data unfolds. This leads to the idea of a class, i.e., a subset of processes defining the constraints that the process must satisfy. However, Olszewski and Sandroni (2009a) extend the basic impossibility theorems to the case where processes are restricted to a closed and convex class

* Corresponding author.
E-mail address: sandroni@kellogg.northwestern.edu (A. Sandroni).

0165-4896/$ – see front matter
doi:10.1016/j.mathsocsci.2012.07.002

(e.g., exchangeable stochastic processes). So, even if the experts are required to deliver a process that belongs to a closed, convex class, Alice will not find a screening contract.

In contrast, Al-Najjar et al. (2010) constructed a screening contract (in fact, a screening test) on a convex class that is not closed. Their class is the set of asymptotically reverse mixing processes that Jackson et al. (1999) characterized, in a De Finetti-style result, as all mixtures of a set of basic processes. Hence, the class satisfies the convexity property.

The positive results in Al-Najjar et al. (2010) and Olszewski and Sandroni (2009b) assume that Bob does not discount the future. In contrast, we consider the case where Bob discounts the future, although he may be as patient as possible (i.e., his discount factor may be arbitrarily close to one, but not one). We show that the contracts in Al-Najjar et al. (2010) and Olszewski and Sandroni (2009b) will not screen informed and uninformed experts if Bob discounts the future even though they can screen informed and uninformed experts in the undiscounted case. Hence, the idea of relaxing technical conditions such as closeness on the class of permitted theories in order to produce a screening contract does not succeed if the future is discounted.

The interest in the discounted case comes from the idea that it may not be helpful for Alice to dismiss the uninformed expert on an arbitrarily remote future date. If Bob discounts the future then he does not fear of punishment in the far distant future. Thus, if Bob is uninformed and discounts the future then he will not undertake Alice's contract only if his process can be dismissed by bounded data sets.

A class is near convex if whenever two processes belong to the class then a mixture of them, with equal weights, also belongs to the class. Clearly, a near convex class can accommodate all classes that can be characterized in a De Finetti-style result (i.e., any class of processes that can be characterized as convex combinations of elementary processes; well-known examples of such convex classes include exchangeable processes, asymptotically reverse mixing processes, and the set of all probability measures with finite support).[1] We show that if a perfectly informed expert accepts Alice's contract then a completely uninformed expert also accepts it. This result holds for any contract provided that the class of permissible theories is near convex and the experts discount the future. Hence, as long as experts discount the future, the impossibility of screening informed and uninformed experts holds as long as theories are restricted in any way such that a mixture of two allowed theories, with equal weight, is also allowed.

Our convexity assumption can be motivated in different ways. First, as mentioned above, this assumption is satisfied in the classes of processes considered in Al-Najjar et al. (2010) and so this result reveals that the screening ability of these procedures requires arbitrarily long data sets. As mentioned above, *any* class of processes that can be characterized in a De Finetti-style result satisfies the convexity property. Finally, convexity can be motivated as follows: suppose that Alice knows that she may not have sufficient data to produce a conventional estimator. So, she can make use of an expert's prior information. That is, if an expert delivers a reliable prior to Alice then she can use the data to update the given prior by Bayes' rule. Examples where an expert's prior is elicited can be found in O'Hagan et al. (2006) and include applications to the effect of nuclear waste on temperature, drug testing, sales of engines and future earnings. However, the space of priors is typically convex. A mixture of two priors is still a prior. Therefore, our convexity proviso (and, hence, our results) holds. It follows that, even in relatively simple problems, Alice cannot determine whether the experts' prior is based on genuine prior information or if it is based on no information at all because her adverse selected problem is not eased by contracts: There is no screening contract that an informed expert who has prior information accepts and an uninformed self-proclaimed expert who has no prior information rejects.

We conclude this introduction by pointing out that our results are false without the near convexity assumption. For instance, the class of all independent, identically distributed processes is not near convex. If Bob is restricted to this class then he must announce a fixed probability at period zero. The uniform law of large numbers shows that there is a future period such that frequencies and data generating probabilities must be close. However, an uninformed expert has no way to ensure that his probability, announced once and for all at period zero, will necessarily match future empirical frequencies.[2] Our results hold if we allow for mixtures of independent, identically distributed processes. As is well-known, this is the class of exchangeable processes. Thus, (near) convexity is a key condition for the impossibility results.

This paper is organized as follows. In Section 2, the model and basic definitions are presented. Section 3 contains the main result. Section 4 concludes and all proofs are in the appendix.

2. Basic definitions

Each period, one outcome, out of a finite set S, is observed. Let S^t denote the Cartesian product of t copies of S, let $\overline{S} = \bigcup_{t\geq 0} S^t$ be the set of all finite histories.[3] We also define $\Omega = S^\infty$ as the set of *paths*, i.e., infinite histories.

Let a *cylinder* with base on $s_m \in S^m$ be the set of all paths such that the first m elements are s_m. We endow Ω with the smallest σ-algebra that contains all such cylinders. We also endow Ω with the product topology (the topology that comprises of unions of cylinders).

Given a set X, we define $\Lambda(X)$ as the set of probability measures on X with finite support. So, an element of $\Lambda(X)$ is a probability measure over a finite subset of X. If X is endowed with a σ-algebra and a topology (or if X is finite) then $\Delta(X)$ is the set of probability measures on X endowed with the weak*-topology.[4]

A function

$$f : \overline{S} \longrightarrow \Delta(S)$$

takes data (finite outcomes) as input and returns a probabilistic forecast for the following period as output. It is well known that f defines a probability measure $P \in \Delta(\Omega)$ on the space of paths. To simplify the language, a probability measure P in $\Delta(\Omega)$ is called a *process*.

Let $\Theta \subseteq \Delta(S)$ be an arbitrary set of processes. We refer to Θ as a *class*and to a process in the class Θ as an *acceptable* process. Before any data is observed, Alice and Bob are convinced that the data generating process belongs to Θ. However, Bob claims that he

[1] Naturally, if theories are not restricted then the near convexity assumption holds because the space of all theories is convex. So, the screening tests of Dekel and Feinberg (2006) and Olszewski and Sandroni (2009b, 2011) also require large data sets and fail to screen informed and uniformed experts when the future is discounted. This special case of our results is, however, known and follows from the results in Sandroni (2003).

[2] See Olszewski and Sandroni (2009a) for a formal exposition on this point.

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

[4] The weak*-topology consists of all unions of finite intersections of sets of the form

$$\left\{Q \in \Delta(X) : \left|E^P h - E^Q h\right| < \varepsilon\right\},$$

where $P \in \Delta(X)$, $\varepsilon > 0$, and h is a real-valued and continuous function on X. See Rudin (1973).

knows more. Bob claims that he knows exactly each process in the class Θ Nature has selected to produce the data. We define

Definition 1. A class $\Theta \subseteq \Delta(\Omega)$ is near convex if

$$P_1 \in \Theta \quad \text{and} \quad P_2 \in \Theta \quad \text{then } 0.5P_1 + 0.5P_2 \in \Theta.$$

So, if processes P_1 and P_2 are acceptable then the process $0.5P_1 + 0.5P_2$ obtained as a mixture of P_1 and P_2 (with equal weights) is also acceptable.

Naturally, the set of all processes is an example of a convex class. There are also other convex classes of interest. For example, a fundamental class of theories is the set of exchangeable processes. As shown by De Finetti (1937), an exchangeable processes is a mixture of independent, identically distributed processes. Hence, the set of exchangeable processes is a convex class. A similar characterization shows that partially exchangeable processes are mixtures of Markov processes (see Diaconis and Freedman, 1980). Hence, the set of partially exchangeable processes is also a convex class. The De Finetti-style result of Jackson et al. (1999) shows that asymptotically reverse mixing processes form another convex class. We refer the reader to Diaconis and Freedman (1987) for a list of De Finetti-style results. Finally, Shmaya and Hu (2010) consider a class of computable distributions that is not convex, but it is near convex.[5]

An important feature in the classes of processes that we consider is the absence of topological assumptions (e.g., compactness) imposed on them. However, our (near) convexity assumption is restrictive and rules out several classes of interest such the class of all independent, identically distributed processes.

2.1. Contracts

Bob may be an informed expert who knows the data generating process. Alternatively, Bob may be an uninformed expert who knows as much as Alice about the data generating process (i.e., he only knows that the data generating process belongs to Θ). Alice wants a contract that Bob, if informed, accepts and Bob, if uninformed, rejects. Bob decides whether to accept Alice's contract at period zero (before any data is observed). If Bob does not accept a contract then he does not deliver a process and his payoff is zero.

Bob must pay a small cost c to deliver his process. Alice does not observe c, but she believes that c is obtained as a draw from a random variable $\tilde{c}$ that has a density and support on $[0, \bar{c}]$, where $\bar{c}$ can be arbitrarily small.

If Bob accepts Alice's contract then he delivers an acceptable process $P \in \Theta$ to Alice at period zero. An initial transfer may occur at period zero (after the process is announced). This transfer gives utility $u(P, \varnothing)$ to Bob. At period t, if data $s_t \in S^t$ is observed a new transfer may occur. Bob's payoff for this contingent transfer is $u(P, s_t)$. We assume that Bob's contingent payoffs are uniformly bounded. i.e., for some $M < \infty$

$$|u(P, s_t)| < M \quad \text{for all } P \in \Theta \text{ and } s_t \in \bar{S}.$$

Bob's discount factor is $\beta < 1$. So, given a path $s \in \Omega$, Bob's discounted net payoff at period zero is

$$U(P, s) = u(P, \varnothing) + \sum_{t=1}^{\infty} \beta^t u(P, s_t), \quad \text{where } s = (s_t, \ldots). \tag{2.1}$$

A *contract* is defined by the payoff function U. We also define $U_P : \Omega \longrightarrow \Re$ as the function such that $U_P(s) = U(P, s)$. If informed, Bob faces common risk and can determine the odds of receiving each contingent payoff. In particular, if Bob believes that the data generating process is $P \in \Theta$ then his expected payoff is

$$E^P\{U_P\} - c,$$

where E^P is the expectation operator associated with $P \in \Theta$.

Definition 2. An informed expert accepts contract U if for any acceptable process $P \in \Theta$ and for any cost $c \in [0, \bar{c}]$,

$$E^P\{U_P\} - c \geq 0. \tag{2.2}$$

Alice knows that Bob, if informed, accepts this contract because Bob's expect utility is positive for all realizations of his private information (i.e., the data generating process and the disutility c of announcing a process).

Now assume that Bob only knows that the data generating process is in Θ. Then Bob faces uncertainty. By definition, Bob does not know which probability measure $P \in \Theta$ he must use to compute the odds of his contingent payoffs. The same problem persists if Bob selects his process by randomizing once (at period zero) according to a random generator of processes $\zeta \in \Lambda(\Theta)$, with finite support. Then, if the data generating process is $P \in \Theta$, Bob's expected payoff is

$$E^{\zeta} E^P U - c = \sum_{i=1}^{n} \gamma_i E^P U_{Q_i} - c,$$

where ζ selects process Q_i with probability γ_i, $i = 1, \ldots, n$.

We assume that Bob evaluates his prospects based on the process $P \in \Theta$ that gives him minimal expected utility. Hence, Bob only accepts the contract in the extreme case that his expected utility is positive, no matter which permissible process generates the data.[6] We purposely consider this extreme decision rule because it will show that our result holds for all other, less extreme decision rules. Formally, Bob's net payoff is

$$\inf_{P \in \Theta} E^{\zeta} E^P U - c,$$

where E^{ζ} is the expectation operator associated with $\zeta \in \Lambda(\Theta)$.

Definition 3. An uninformed expert accepts contract U if for any $\delta > 0$ there exists a random generator of theories $\bar{\zeta} \in \Lambda(\Theta)$ such that for all costs $c \in [0, \bar{c} - \delta]$,

$$\inf_{P \in \Theta} E^{\bar{\zeta}} E^P U - c \geq 0. \tag{2.3}$$

That is, the uninformed expert accepts the contract if there exists a random generator of theories $\bar{\zeta} \in \Lambda(\Theta)$ such that his expected utility is positive, *no matter which process in Θ actually generates the data.* This holds for all costs of announcing a process except for those above $\bar{c} - \delta$ (the odds that Bob's cost is above $\bar{c} - \delta$ can be made arbitrarily small, if δ is small enough). A contract *U screens informed and uninformed experts* if the informed expert accepts the contract, but the uninformed expert does not accept it.

3. Main result

Proposition 1. *Let Θ be a near convex class. If an informed expert accepts a contract U then an uninformed expert also accepts the contract U.*

[5] Unlike this paper, Shmaya and Hu (2010) consider computational constraints that the expert might have. Our results indicate that computational constraints, or some other limitation on the expert's capabilities, are indeed essential conditions if a screening method is to be constructed on a near-convex class of processes.

[6] See Gilboa and Schmeidler (1989) and Gajdos et al. (2008), among several contributions, for axiomatized behavioral rules under uncertainty.

Proposition 1 shows that no contract can screen informed and uninformed experts. So, a contract does not mitigate Alice's adverse selection problem. This holds even though Alice is willing to pay large amounts to learn the relevant probabilities, she will have data at her disposal to assess the performance of Bob's process and she can stipulate large (though not unboundedly large) contingent penalties and rewards. Moreover, Bob, if uninformed, only accepts the contract if he is assured to receive positive payoff, no matter which permissible process runs the data.

The key restrictions in Proposition 1 are that the expert discounts the future and the near convexity property of the set of acceptable processes. As we mentioned, the discount factor can be arbitrarily close to one and Alice may have at her disposal long strings of data. However, the relevant size of the data set is bounded in the sense that the assumption of a discounted future rules out the possibility of having increasing long and relevant strings of data for a sequence of processes within the permissible class. As mentioned above, convexity is satisfied in classes of interest such as exchangeable processes, partially exchangeable processes, asymptotically reverse mixing processes, or any class of process for which a De Finetti-style result can be obtained (i.e., the class of processes can be characterized as mixtures over a set of basic processes). It is also satisfied in the case that the expert must announce a prior belief reflecting prior information about the relevant stochastic processes. Thus, it is not possible to determine whether the expert's prior reflects genuine information about the relevant process.

We conclude this section with an informal description of a simple corollary of our main result. Assume that the self-proclaimed expert makes the weaker claim that the data generating process belongs to some finite set of processes in Θ. It is straightforward to modify our proof and demonstrate that it is impossible to screen informed and uninformed experts in the case of a discounted future. This result holds for an *arbitrary* class of process Θ. In particular, Θ need not be convex or near-convex. A proof can be found from the authors upon request.

3.1. Intuition behind the proof

The proof of Proposition 1 uses the minmax theorem which is a basic tool in this literature. The basic idea is to set up a game between Nature and the uninformed expert, Nature picks a process in the class and the experts also picks a process in the class (perhaps at random). Hence, if the minmax theorem holds then no matter what process Nature actually selects, the expert's payoff is at least as high as what he could get if he knew Nature's strategy. The main difficulty here is that for the minmax theorem to hold we need the strategy space of one player to be compact. Given that we do not assume that the class is closed this does not hold.

A way around this difficulty is to let Nature select processes from the closure of the class and to show that the expert (if he knew Nature's strategy) could choose a process in the class (not necessarily in the closure of the class) that would a yield sufficiently high payoff. It is here that the assumption of a discounted future is essential. This is not necessarily true if the future is not discounted.

4. Conclusion

In the case of a discounted future, the impossibility results showing that contracts cannot screen informed and uninformed experts can be extended to the case of where theories are restricted to a near convex class. Hence, convexity seems to be the crucial condition supporting the impossibility of screening informed and uninformed experts. In particular, even in relatively simple problems, it is not possible to determine whether an expert's announced prior is based on relevant prior information.

5. Proofs

Recall that a *cylinder* with base on $s_t \in \{0,1\}^t$ is the set $C(s_t) \subset \{0,1\}^\infty$ of all paths such that s_t are their first t elements. Also recall that $\Delta(\Omega)$ is endowed with the weak-$*$ topology and is metric space. If $P^n \in \Delta(\Omega)$ converges to $P \in \Delta(\Omega)$ as n goes to infinity then for every $s_t \in \{0,1\}^t$,

$$P^n(C(s_t)) \xrightarrow[n\to\infty]{} P^n(C(s_t)).$$

Given a probability measure $Q \in \Delta(\Omega)$ let E^Q be the expectation operator associated with Q. Let N be the set of natural numbers.

Lemma 1. *Fix $\varepsilon > 0$. Assume that an informed expert accepts a contract U. Also assume that $P^n \in \Theta$ converges to $P \in \Delta(\Omega)$ as n goes to infinity. Then, for some $\bar{n} \in N$*

$$E^P U_{P^{\bar{n}}} \geq \bar{c} - \varepsilon.$$

Proof. By assumption,

$$E^{P^n} U_{P^n} \geq \bar{c}.$$

Also,

$$\begin{aligned} E^P U_{P^n} &= E^{P^n} U_{P^n} + E^P U_{P^n} - E^{P^n} U_{P^n} \\ &\geq \bar{c} - \left| E^P U_{P^n} - E^{P^n} U_{P^n} \right|. \end{aligned}$$

In addition, given that $\beta < 1$ and u is uniformly bounded, there exists $m \in N$ such that

$$\begin{aligned} U_{P^n}(s) &= u(P^n, \varnothing) + \sum_{t=1}^{m} \beta^t u(P^n, s_t) \\ &\quad + \sum_{t=m+1}^{\infty} \beta^t u(P^n, s_t) \quad \text{where } s = (s_t, \ldots) \end{aligned}$$

and

$$\left| \sum_{t=m+1}^{\infty} \beta^t u(P^n, s_t) \right| \leq \frac{\varepsilon}{2} \quad \text{for all } s_t.$$

So,

$$U_{P^n}(s) = U^1_{P^n}(s) + U^2_{P^n}(s),$$

where

$$U^1_{P^n}(s) = u(P^n, \varnothing) + \sum_{t=1}^{m} \beta^t u(P^n, s_t);$$

$$U^2_{P^n}(s) = \sum_{t=m+1}^{\infty} \beta^t u(P^n, s_t), \quad s = (s_t, \ldots).$$

By construction,

$$\left| E^P U^2_{P^n} - E^{P^n} U^2_{P^n} \right| \leq \frac{\varepsilon}{2}.$$

Let $k = \sum_{i=0}^{m} 2^i$. Then, there exists $\bar{n}$ such that

$$\left| P^{\bar{n}}(C(s_t)) - P^{\bar{n}}(C(s_t)) \right| \leq \frac{\varepsilon}{2Mk} \quad \text{for all } s_t.$$

So,

$$\left| E^P U^1_{P^{\bar{n}}} - E^{P^n} U^1_{P^{\bar{n}}} \right| \leq \frac{\varepsilon}{2}.$$

It follows that

$$\left| E^P U_{P^{\bar{n}}} - E^{P^n} U_{P^{\bar{n}}} \right| \leq \varepsilon. \quad \square$$

Let $\bar{\Theta}$ be the closure of Θ. Let the function $H : \bar{\Theta} \times \Lambda(\Theta) \to R$ be defined by

$$H(P, \zeta) := E^{\zeta} E^{P} U.$$

Lemma 2. *H is a continuous function of P.*

Proof. By definition,

$$H(P, \zeta) := \sum_{i=1}^{n} \gamma_i E^{P} U_{Q_i},$$

where ζ selects process Q_i with probability γ_i, $i = 1, \ldots, n$. Hence, it suffices to show that for a fixed $Q_i \in \Delta(\Omega)$, $E^P U_{Q_i}$ is a continuous function of P. Given that $\Delta(\Omega)$ (and, hence, $\bar{\Theta}$) is endowed with the weak-$*$ topology it suffices to show that U_{Q_i} is a continuous function of s.

Assume that $s^n \in \Omega$ converges to $s \in \Omega$ as n goes to infinity. Fix $x > 0$. Let m be large enough so that

$$\left| \sum_{t=m+1}^{\infty} \beta^t u(Q_i, s_t) \right| \leq \frac{x}{4} \quad \text{for all } s_t.$$

By definition,

$$U_{Q_i}(s^n) = u(Q_i, \varnothing) + \sum_{t=1}^{m} \beta^t u(Q_i, s_t^n) + \sum_{t=m+1}^{\infty} \beta^t u(Q_i, s_t^n) \quad \text{where } s^n = (s_t^n, \ldots)$$

and

$$U_{Q_i}(s) = u(Q_i, \varnothing) + \sum_{t=1}^{m} \beta^t u(Q_i, s_t) + \sum_{t=m+1}^{\infty} \beta^t u(Q_i, s_t) \quad \text{where } s = (s_t, \ldots).$$

So,

$$\left| U_{Q_i}(s^n) - U_{Q_i}(s) \right| \leq \left| \sum_{t=1}^{m} \beta^t \left(u(Q_i, s_t^n) - u(Q_i, s_t) \right) \right| + \left| \sum_{t=m+1}^{\infty} \beta^t u(Q_i, s_t^n) \right| + \left| \sum_{t=m+1}^{\infty} \beta^t u(Q_i, s_t^n) \right|. \quad (5.1)$$

If n is sufficiently large then the first term in (5.1) is zero. Hence, $\left| U_{Q_i}(s^n) - U_{Q_i}(s) \right|$ is less than $\frac{x}{2}$ if n is sufficiently large. $\square$

Lemma 3. *If Θ is near convex then $\bar{\Theta}$ is convex.*

Proof. Let $P^1 \in \bar{\Theta}$, $P^2 \in \bar{\Theta}$, and $\lambda \in [0, 1]$. By definition, there are sequences $\{P_n^1\}$ and $\{P_n^2\}$ such that for every n, $P_n^1 \in \Theta$, $P_n^2 \in \Theta$, and for every cylinder C, with base on a finite history, $P_n^1(C) \longrightarrow P^1(C)$ and $P_n^2(C) \longrightarrow P^2(C)$ as n goes to infinity. Let $\{\lambda_n\}$ be a sequence of rational numbers that converge to λ as n goes to infinity. Then, for every n, $\lambda_n P_n^1 + (1 - \lambda_n) P_n^2 \in \Theta$ and for every cylinder C,

$$(\lambda_n P_n^1 + (1 - \lambda_n) P_n^2)(C) \longrightarrow (\lambda P^1 + (1 - \lambda) P^2)(C)$$

as n goes to infinity. Thus, $(\lambda P^1 + (1 - \lambda) P^2) \in \bar{\Theta}$. $\square$

Theorem 5.1 (Fan, 1953). *Let X be a compact Hausdorff space, which is a convex subset of a linear space, and let Y be a convex subset of linear space (not necessarily topologized).*[7] *Let H be a real-valued function on $X \times Y$ such that for every $y \in Y$, $H(x, y)$ is lower semi-continuous with respect to x. If H is also convex with respect to x and concave with respect to y (for every $y \in Y$ and for every $x \in X$, respectively), then*

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

[7] Fan allows for X and Y that may not be subsets of linear spaces. We, however, apply his theorem only to subsets of linear spaces.

Proof of Proposition 1. Fix $\delta > 0$. It is immediate that H is a linear function on both arguments. By Lemma 2, H is a continuous function of P. By the Riesz and Banach–Alaoglu Theorems, $\Delta(\Omega)$ is a compact space in weak-$*$ topology; it is a metric space, and so Hausdorff, (see for example Rudin, 1973, Theorem 3.17). So, $\bar{\Theta}$ (the closure of Θ) is a compact subset of $\Delta(\Omega)$. By Lemma 3, $\bar{\Theta}$ is convex.

Hence, by Fan's (1953) Minimax Theorem,

$$\min_{\bar{\Theta}} \sup_{\Lambda(\Theta)} H = \sup_{\Lambda(\Theta)} \min_{\bar{\Theta}} H.$$

By Lemma 1, for any $P \in \bar{\Theta}$ there exists $P' \in \Theta$ such that

$$E^P U_{P'} \geq \bar{c} - \frac{\delta}{2}.$$

Let $\zeta \in \Lambda(\Theta)$ be the Dirac-measure on Θ that puts all mass on P'. Then, $H(\zeta, P) \geqslant \bar{c} - \frac{\delta}{2}$. So,

$$\min_{\bar{\Theta}} \sup_{\Lambda(\Theta)} H \geqslant \bar{c} - \frac{\delta}{2} \Rightarrow \sup_{\Lambda(\Theta)} \min_{\bar{\Theta}} H \geqslant \bar{c} - \frac{\delta}{2}.$$

Hence, there exists $\zeta \in \Lambda(\Theta)$ such that $H(P, \zeta) \geqslant \bar{c} - \delta$ for every $P \in \bar{\Theta}$ (and, in particular, for every $P \in \Theta$). $\square$

Acknowledgments

This paper was previously titled "A Difficulty in the Testing of Theories". We thank the Editor and two anonymous referees for useful comments. Sandroni gratefully acknowledges financial support from the National Science Foundation. All errors are ours.

References

Al-Najjar, N., Smorodinsky, R., Sandroni, A., Weinstein, J., 2010. Testing theories with learnable and predictive representations. Journal of Economic Theory 145 (6), 2203–2217.

De Finetti, B., 1937. La prévision: ses lois logiques, ses sources subjectives. Annales de l'institut Henri Poincaré 7, 1–66.

Dekel, E., Feinberg, Y., 2006. Non-Bayesian testing of a stochastic prediction. Review of Economic Studies 73, 893–906.

Diaconis, P., Freedman, D., 1980. De Finetti's theorem for Markov chains. The Annals of Probability 8, 115–130.

Diaconis, P., Freedman, D., 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, K., 1953. Minimax theorems. Proceedings of the National Academy of Sciences of the United States of America 39, 42–47.

Gajdos, T., Hayashi, T., Tallon, J-M., Vergnaud, J-C., 2008. Attitude toward imprecise information. Journal of Economic Theory 140 (1), 27–65.

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with a non-unique prior. Journal of Mathematical Economics 18, 141–153.

Jackson, M., Kalai, E., Smorodinsky, R., 1999. Bayesian representation of stochastic processes under learning: de Finetti revisited. Econometrica 67, 875–893.

O'Hagan, A., Buck, C., Daneshkan, A., Eiser, J., Garthwaite, P., Jenkinson, D., Oakely, J., Rakow, T., 2006. Uncertain Judgement: Eliciting Experts' Probabilities. John Wiley & Sons.

Olszewski, W., Sandroni, A., 2007. Contracts and uncertainty. Theoretical Economics 2, 1–13.

Olszewski, W., Sandroni, A., 2009a. Strategic manipulation of empirical tests. Mathematics of Operations Research 34, 57–70.

Olszewski, W., Sandroni, A., 2009b. A nonmanipulable test. Annals of Statistics 37 (2), 1013–1039.

Olszewski, W., Sandroni, A., 2011. Falsifiability. American Economic Review 101 (2), 788–818.

Rudin, W., 1973. Functional Analysis. McGraw-Hill, Inc.

Sandroni, A., 2003. The reproducible properties of correct forecasts. International Journal of Game Theory 32, 151–159.

Shmaya, E., Hu, T., 2010. Expressible Inspections, Working paper, Northwestern University.