DUAL SCORING

CHRISTOPHER P. CHAMBERS*, PAUL J. HEALY**, AND NICOLAS LAMBERT***

ABSTRACT. For any strictly proper scoring rule, we provide a full dual characterization of the optimal announcement of an agent with a quasiconcave, continuous, and increasing utility function. The dual characterization leverages the notion of indirect utility. We use this characterization to construct a strictly proper scoring rule which uniformly bounds misreports when risk aversion is bounded above by a known threshold. Further, for any strictly proper scoring rule and tolerance for error, misreports are uniformly bounded above by the tolerance for some nontrivial risk aversion parameter.

Keywords: Scoring rule; probability forecast; general preferences.

JEL Classification: D81, D84, C90.

I. INTRODUCTION

A scoring rule is an incentive device for obtaining probabilistic assessments from agents in the face of subjective uncertainty. Formally, it can be thought of as a menu of state-contingent monetary payoffs, indexed by the set of possible beliefs of the agent, from which the agent is asked to choose. If the scoring rule is proper, then a risk-neutral agent maximizes her expected utiltiy by choosing from the menu the state-contingent payoff that is associated with her subjective belief.[1]

Though there is a long tradition of using scoring rules in experiments (for example, Nyarko and Schotter, 2002 use a scoring rule to elicit beliefs in games), the assumption of risk neutral expected utility maximization is of course suspect. Thus, we study behavior in scoring rules for agents who may be risk averse, or may not even be expected utility maximizers. This includes agents whose decision-making is not consistent with probabilistic sophistication (Machina and Schmeidler, 1992).[2] An important question is

*Dept. of Economics, University of California San Diego, 9500 Gilman Drive, La Jolla, CA, USA.
**Dept. of Economics, The Ohio State University, 1945 North High Street, Columbus, OH, USA.
**Graduate School of Business, Stanford University, 655 Knight Way, Stanford, CA, USA. .

[1]This interpretation is firmly in line with the revealed preference tradition, whereby an agent may not necessarily perceive that she holds a probabilistic assessment, but nevertheless her behavior is consistent with such a belief. To this end, we need not ask the agent to "report" a belief, but rather to choose from a set. It is usually easier to use the language of "reporting" beliefs, however. There is no real reason to require that there be a unique state-contingent payoff associated with each probability, but it is conventional to do so, and not much is gained by weakening this assumption.

[2]Although such agents do not "hold probabilities", they can still report a probability when faced with a scoring rule, and their report can still be useful for making inferences about their preferences.

how such agents optimize when facing a given scoring rule, and to what extent their reports may differ from their true belief when they are probabilistically sophisticated.

In this paper, we provide a full characterization of optimal behavior for individuals with "standard" economic utility functions over state-contingent payoffs, which are quasiconcave, continuous, and strictly increasing. This includes agents with expected utility preferences (with or without risk neutrality), but also ambiguity averse agents who do not adhere to any concept of "likelihood" or probability. Indeed, the goal is to study choice behavior in scoring rules without assuming any particular decision-theoretic model of choice.

A direct precedent to this paper is Grünwald and Dawid (2004), which describes two classical approaches to the problem of robust statistics. Given is a convex and compact set of probability measures; and the goal is to "select" a probability from this set. Assume that state-contingent monetary payoffs are evaluated according to the minimal expected value according to all probabilities in the set. Using minimax theory, Grünwald and Dawid obtain a basic duality characterization: Such an individual facing a proper scoring rule can be viewed *as if* she is minimizing a strictly convex function on the set of probabilities. Each proper scoring rule is associated with its own strictly convex function. For example, the authors observe the duality between the logarithmic scoring rule (Good, 1952) and the entropy function.[3] The result in this note directly generalizes Grünwald and Dawid (2004) in an environment with finite states.[4]

The duality we study is the duality between a *direct* utility function and an *indirect* utility function. There are several related such dualities, but we focus on one which has recently been exploited fruitfully by Cerreia-Vioglio et al. (2011b).[5] For a given utility function U—usually understood as a "direct" utility over state-contingent payoffs—we define an "indirect utility" over price-wealth pairs in the natural way:

$$G(p,w) = \sup\{U(x) : p \cdot x \leq w\}.$$

That is, $G(p,w)$ asserts the maximal utility achievable by an individual with utility U when "market prices" are p, and the wealth available for expenditure is w. If p

[3]See also Chambers (2008), who independently proved the same result in a much simpler environment.
[4]Grünwald and Dawid (2004) includes other results as well.
[5]This is arguably the "right" duality notion to use in our case, where we want to allow scoring rules to potentially pay negative monetary amounts. These notions have existed in economics at least since Konüs (1939) and Ville and Newman (1952), which are translations of earlier foreign language works, whereas Roy (1947) arguably popularized the concept. de Finetti (1949) established an early duality result using such functions. The related duality studied by Shepherd (1970), Lau (1969), Cornes (1992), Weymark (1980) requires all payoffs to be nonnegative.

are interpreted as probabilities, then equivalently, the indirect utility gives the highest utility achievable to U by a state-contingent payoff with expectation at most w.[6]

Our main result is that, for any decision maker who has a quasiconcave, weakly increasing, and continuous utility function over state-contingent payments, the unique optimal announcement p in scoring rule f coincides with the unique p which minimizes $G(p, V(p))$, where $V(p) = p \cdot f(p)$ is the "value function" associated with f that gives the expected payoff of announcing p.

As we see it, there are four main reasons why such a duality result is interesting.

- First and foremost, working with the dual problem allows one to derive new results of importance with simple proofs; results that would otherwise be challenging to obtain via a direct approach. For example, Theorem 3 below establishes that if risk aversion is not "too high", then we can uniformly bound the magnitude of misreports. Although other authors have shown that misreports can be bounded by "flattening" the scoring rule, using the duality result allows us to verify more easily that they can be *uniformly* bounded.
- Second, many preference specifications in economics are defined only via their indirect utility functions. Chief among these preference classes is the *Gorman polar form* (Gorman, 1961), described below in detail and used heavily in applied modeling.
- Third, though we don't explore it here, duality results are generally quite useful in comparative static exercises. Results from Cerreia-Vioglio et al. (2011b) can be used to study comparative notions of risk aversion (for example, that of Yaari (1969)), and how subjects behavior when facing different scoring rules changes when becoming more risk averse.
- Fourth, assuming a finite state space, our result generalizes that of Grünwald and Dawid (2004) to a broad class of preferences over uncertain prospects.

To be clear, our main theorem itself does not provide new tools or methods for practitioners. Rather, we provide a mathematical framework that may make certain analyses of scoring rules more tractable, which in turn can lead to new innovations. For example, we use our main result to establish that we can elicit, to an arbitrarily high degree of precision, an individual's subjective probability when their preferences are in the constant absolute risk aversion (CARA) class and their risk-aversion is known to be below

[6]There is a sense in which the indirect utility is related to the notion of Fenchel conjugation. To see this, observe that the Fenchel conjugate of a monotonic U can be written as: $G(y) = \inf_x x \cdot y - U(x)$, for $y \geq 0$ (not necessarily a probability). We can think of the negation of this problem, $-G(y) = \sup_x U(x) - x \cdot y$. The value of this problem is clearly equivalent to the value of the problem $\sup_{(x,t)} U(x) + t$ subject to the constraint that $t + x \cdot y \leq 0$. This is a "cardinal" version of the indirect utility when there is a "numeraire" good t with a fixed price of 1, and when total wealth is zero. When the numeraire good has a fixed price of 1, we cannot renormalize prices to sum to 1.

some finite bound. Similar insights are found in Kadane and Winkler (1988) and Armantier and Treich (2013) for the quadratic scoring rule; our result differs because it generalizes this insight to any proper scoring rule, and also provides a uniform convergence result across all utility indices in the class and all probabilities (as opposed to a convergence result for a fixed utility index and probability belief). As far as we know, this is the first result of this type in the literature. The duality approach may also be useful in proving a host of other new results which we have not explored here.

There are many examples of utility specifications where the duality is especially simple. One such example is the case of ambiguity-averse agents with multiple priors studied by Grünwald and Dawid (2004). Another is the class of *translation invariant* utility functions. These are utility functions for which wealth effects are absent. This is operationalized by assuming that adding a dollar in each state of the world translates into an additional unit of utility. In this case, the dual minimization problem takes the form of minimizing the sum of the value function, and some convex function of probabilities, specific to the utility in question. A particular special case of translation invariant preferences is the family of CARA preferences. These are the unique subjective expected utility preferences which can be expressed in a translation invariant form. The convex function in this case is the *relative entropy* function, relative to the subjective probability in this case. Bickel (2007) describes the optimization problem for such individuals facing scoring rules.

A further generalization of translation invariant preferences is provided by the Gorman polar form preferences (Gorman, 1961). In our context, these are preferences for which there is some $\beta \geq 0$, $\beta \neq 0$ for which adding t units of β to consumption adds t units of utility. These preferences are highly useful in applied modeling, as they allow one to meaningfully describe a "group" of individuals as a single individual, behaving in her own best interest.

Finally, to see why this directly generalizes Grünwald and Dawid (2004), observe that when $U(x) = \min_{p \in P} p \cdot x$, then $G(p,w) = w$ when $p \in P$ and $+\infty$ otherwise. In other words, the uniquely optimal announcement for such a utility function is equivalently given by $\min_{p \in P} V(p)$. While Grünwald and Dawid (2004) and Chambers (2008) each rely on versions of the minimax theorem, the proof here is simpler and directly leverages the separating hyperplane theorem.[7]

Section II presents our main result, as well as a related result stemming from Roy's identity. Section IV presents several examples, illustrating how our result can be used. Section V describes a method of bounding misreports to any arbitrarily high degree of precision, starting from any proper scoring rule. Finally, section VI concludes.

[7]A proof can also be established using minimax.

II. Optimal values for general decision models

Let Ω be a finite set of *states* and $\Delta(\Omega)$ the set of probability measures on Ω. A *scoring rule* is a mapping $f : \Delta(\Omega) \to \mathbb{R}^{\Omega}$. It is *proper* if for all $p, p' \in \Delta(\Omega)$, $p \cdot f(p) \geq p \cdot f(p')$ and *strictly proper* if for all $p, p' \in \Delta(\Omega)$ for which $p \neq p'$, we have $p \cdot f(p) > p \cdot f(p')$. For a strictly proper scoring rule f, define the associated *value function* $V : \Delta(\Omega) \to \mathbb{R}$ by $V(p) = p \cdot f(p) = \sup_{p' \in \Delta(\Omega)} p \cdot f(p')$.[8]

We assume that preferences over state-contingent payoffs are represented by a utility function $U : \mathbb{R}^{\Omega} \to \mathbb{R}$ that is weakly increasing, quasiconcave, and continuous.[9] We call such a utility *standard*, as these assumptions are the basis for classical demand and general equilibrium theory, for example.[10] The *indirect utility* is defined by $G(p, w) = \sup\{U(x) : p \cdot x \leq w\}$, which takes values in $\mathbb{R} \cup \{+\infty\}$. Quite generally, we have $U(x) = \inf_{p \in \Delta(\Omega)} G(p, p \cdot x)$.[11] This is the duality we use for our results.

We emphasize that the only useful economic content of this model are the properties of increasingness and quasiconcavity of preference. Thus, this model can incorporate decision makers with the following types of utility functions; *i.e.* risk-averse expected utility:

$$U(x) = u^{-1}\left(\sum_{\omega \in \Omega} u(x_\omega) p(\omega)\right),$$

where u is increasing and concave, and p is a probability measure. This is a classical certainty-equivalent representation of a risk-averse expected utility maximizer. Of course, the representation of this preference would involve a complex G function, but there is such a function; in fact, an explicit representation of such functions appears in Cerreia-Vioglio et al. (2011b). However, the class of utility functions we can accommodate is much broader than the risk-averse expected utility class, and includes ambiguity averse agents who are not probabilistically sophisticated. Note that agents can still participate meaningfully in a scoring rule even if they don't "use" probabilities in their own decision calculus, since the reported probabilities p will still lead to meaningful inferences about U.

The following is our main result. It claims that for any standard utility and any strictly incentive compatible scoring rule, there is a unique optimal announcement $p^* \in \Delta(\Omega)$, and further, this unique announcement can be arrived at using dual techniques.

[8] For a general, non-proper scoring rule, one can define the value function via $V(p) = \sup_{p'} p \cdot f(p')$.

[9] Weakly increasing means that if $x_\omega > y_\omega$ for all $\omega \in \Omega$, then $U(x) > U(y)$.

[10] In applications, U is known to the agent but not to the experimenter. Rather, the experimenter observes the reported p and makes an inference about U.

[11] This duality is related to, but distinct from, the duality presented in Shepherd (1970), Lau (1969), Cornes (1992); and Weymark (1980). That notion of duality is for functions on the nonnegative orthant. See Cerreia-Vioglio et al. (2011b) for details.

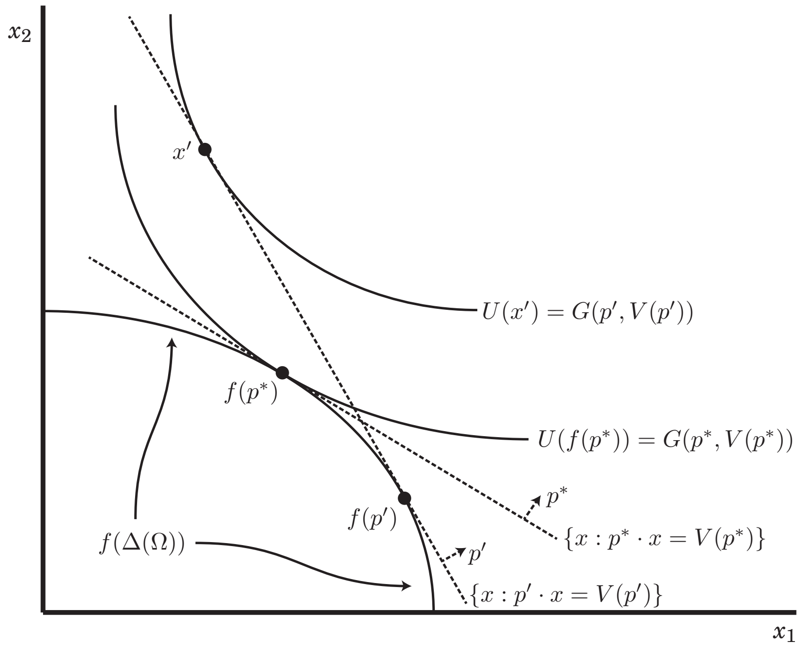

FIGURE I. An illustration of the main theorem.

Namely, it solves the problem $\min_p G(p, V(p))$. The latter problem is often easier to describe, and can allow for richer comparative statics.

Theorem 1. Suppose that f is a continuous and strictly proper scoring rule. Then for a standard utility with indirect utility function G, there is a unique solution to the problem $\max_{p^* \in \Delta(\Omega)} U(f(p^*))$, and it is given by $\arg\min_{p^* \in \Delta(\Omega)} G(p^*, V(p^*))$, where V is the value function associated with f.

The proof of Theorem 1 appears in Appendix A. We sketch the idea of the proof using Figure I. In this diagram Ω contains only two states, so $X = \mathbb{R}^2$. By strict properness, the range of the scoring rule f forms the upper boundary of a strictly convex set. The problem of maximizing U over this set results in an optimal payment vector $f(p^*)$, which is achieved by announcing the distribution p^*.[12] The payment of $f(p^*)$ gives an expected value (under p^*) of $V(p^*) = p^* \cdot f(p^*)$.

Now consider the 'indirect utility' maximization program: For any p we calculate the expectation of the scoring rule under p—which equals $V(p)$—and ask what point $x \in \mathbb{R}^2$ maximizes U subject to the constraint that $p \cdot x \leq V(p)$. This constraint is shown for p^* and p' as dashed lines. For p^* the maximizing point is again $f(p^*)$, which gives indirect utility $G(p^*, V(p^*)) = U(f(p^*))$. But for p' the constraint includes points strictly better than $f(p^*)$ for the decision maker, so the maximum indirect utility (which obtains at

[12] We reiterate that the decision maker need not have 'true' beliefs p^*—or any probabilistic beliefs at all. Here, p^* is interpreted only as the decision maker's optimal announcement given f.

x') is $G(p',V(p')) > U(f(p^*)) = G(p^*,V(p^*))$. By the strictly convex shape of $f(\Delta(\Omega))$ this is true for any $p' \neq p^*$. Thus, the original utility-maximizing point $f(p^*)$ is the unique minimizer of the indirect utility function $G(p,V(p))$.

Theorem 1 can be generalized to include scoring rules which take infinite-valued payoffs (such as the classical logarithmic scoring rule). In this case, existence of an optimal announcement is not guaranteed, but when there is such an announcement, the duality will hold.

The result actually says little more than that the announced p^* separates the convex hull of the image of payoffs of the scoring rule, and the upper contour set of the preference. But it leads to many interesting conclusions when written in this form. For example, by exploiting Roy's identity (Roy, 1947) and the fact that a scoring rule is a subdifferential of its homothetic extension, we get the following:

Corollary 1. Suppose U is standard and f is a continuous and strictly proper scoring rule. If $p^* \in \arg\max_{p\in\Delta(\Omega)} U(f(p))$, we have $f(p^*) \in \arg\max_{\{x:p^*\cdot x \leq V(p^*)\}} U(x)$.

The preceding corollary states that a standard decision maker will choose to announce the p^* at which her Walrasian demand (when given wealth $V(p^*)$) includes $f(p^*)$. In this sense, this result illustrates the connection between the optimal choice of a decision maker from a scoring rule, and the equilibrium price in a Robinson Crusoe economy (the output of the scoring rule playing the role of a technology here).

III. Interpretation

As we see it, as far as economic behavior is concerned, the relevance of the probability distribution pertains to whether choice behavior can be synopsized by a single parameter (the probability, or beliefs). It is well-known that, in general, this cannot be done. To this end, we feel we should offer an interpretation of the optimal p^*.

We are used to thinking as a "preference" as a normative concept. But as a positive concept, and from the revealed preference standpoint, "observing" a preference is problematic. We have remarked on this elsewhere (Azrieli et al. (2012)). Under the standard interpretation of the preference-rationalization model, it is unreasonable to generally expect to uncover a preference relation by offering a sequence of repeated choice experiments. This is especially true in models of choice under uncertainty, where hedging behavior plays a major role. If the preference rationalization model is "true," at best, without adding additional assumptions, we can hope to observe one single choice made from one single budget. This is what we observe with the scoring rule.

Here, we offer an interpretation of Theorem 1, as well as an interpretation of the announced p^*. We start with the following simple observation.[13]

Theorem 2. Let f be continuous and proper. If $p^* \in \Delta(\Omega)$ has the property that for all $\omega, \omega' \in \Omega$, $f(p^*)(\omega) = f(p^*)(\omega')$, then $p^* \in \arg\min_{p \in \Delta(\Omega)} V(p)$. Conversely, if $p^* \in \Delta_{++}(\Omega)$ and $p^* \in \arg\min_{p \in \Delta(\Omega)} V(p)$, then p^* has the property that for all $\omega, \omega' \in \Omega$, $f(p^*)(\omega) = f(p^*)(\omega')$.[14]

Proof. First, suppose that $p^* \in \Delta(\Omega)$ has the property that for all $\omega, \omega' \in \Omega$, $f(p^*)(\omega) = f(p^*)(\omega') = V(p^*)$. Then by definition, for any $p \neq p^*$, it follows $V(p^*) = p \cdot f(p^*) \leq p \cdot f(p) = V(p)$.

Conversely, suppose that $p^* \in \Delta_{++}(\Omega)$ and $p^* \in \arg\min_{p \in \Delta(\Omega)} V(p)$. Observe that by properness, for any p, $0 \leq V(p) - V(p^*) \leq (p - p^*) \cdot f(p)$, so $(p - p^*) \cdot f(p) \geq 0$. Pick any two states, $\omega', \omega'' \in \Omega$, $\omega' \neq \omega''$. Because $p^* \in \Delta_{++}(\Omega)$, there is $\epsilon > 0$ small so that p_ϵ as defined following is an element of $\Delta(\Omega)$: $p_\epsilon(\omega') = p^*(\omega') + \epsilon$, $p_\epsilon(\omega'') = p^*(\omega'') - \epsilon$, and finally $p_\epsilon(\omega) = p^*(\omega)$ for all $\omega \notin \{\omega', \omega''\}$. Observe then that $\epsilon(f(p_\epsilon)(\omega') - f(p_\epsilon)(\omega'')) \geq 0$. Since $\epsilon > 0$ was arbitrary, we conclude by continuity of f that $f(p^*)(\omega') \geq f(p^*)(\omega'')$. Finally, the result follows as ω', ω'' were arbitrary. □

Theorem 2 affords an interpretation of V as being a measure of *dispersion* in payoffs, or more suggestively a measure of uncertainty across states for scoring rule f. When V is minimized, the payoff to the subject from f is state-independent, and there is no uncertainty present. This suggests that the problem $\min_{p \in \Delta(\Omega)} V(p)$ can be interpreted as a problem whereby an individual minimizes the uncertainty across states.[15]

Regarding the indirect utility function, suppose we start with a utility function U which is a *certainty equivalent* representation. Namely, one which has the property that for all $x \in \mathbb{R}$, $U(x, x, \ldots, x) = x$.[16] This function measures utility in monetary units. Then $G(p, w)$ measures the monetary value to the individual with wealth w and who faces state-prices p. Clearly, $G(p, w) \geq w$; and there is (at least one) $p^* \in \Delta(\Omega)$ for which $G(p^*, w) = w$.[17]

The higher is $G(p, w)$, the "further away" the subject is from optimizing at $(w, w, \ldots, w)$. A high value of $G(p, w)$ relative to w reflects a willingness to stake a bet against the odds

[13]In the following, only the second part of the result relies on f being continuous. The first statement is valid more generally.

[14]Here, $\Delta_{++}(\Omega)$ refers to the set of full-support probabilities, or the interior of the simplex.

[15]Theorem 2 can be used to establish relatively easy corollaries on subgradients of V for proper continuous f. For example, for an act $x \in \mathbb{R}^\Omega$ and any $p^* \in \Delta(\Omega)$, if $f(p^*)(\omega) - x_\omega$ is identical across states, then $p^* \in \arg\min_{p \in \Delta(\Omega)} V(p) - p \cdot x$, and conversely if $p^* \in \Delta_{++}(\Omega)$. These ideas can also be seen as an application of McCarthy (1956).

[16]Existence of such a utility function is easy to establish for any preference in our domain.

[17]Details of this are in Cerreia-Vioglio et al. (2011b).

p with wealth w. This willingness to bet may be informed by subjective probabilities and risk attitudes, but is meaningful more generally.

For example, a risk-averse subject with subjective probability p^* is unwilling to bet against the odds of p^*, and hence $G(p^*,w) = w$. On the other hand, highly risk averse subject may hold a subjective probability of p^*, but still may not be willing to bet much against odds $q \neq p^*$ due to risk aversion. She would prefer instead to keep her wealth (relatively) state-independent. In this case, $G(q,w)$ is close to w. A risk-neutral subject with subjective probability p^* will be willing to bet against any odds $p \neq p^*$. Because of risk-neutrality, they will be willing to invest arbitrarily large amounts of money in such a bet. Hence, for such an individual, $G(p,w) = +\infty$. Finally, for a maxmin decision maker, who simply wants to maximize the payoff in the worst state, $G(p,w) = w$. This individual is unwilling to deviate from certainty.

So, the function G as applied to V serves as an "adjustment" for attitudes toward dispersion in payoffs across states. All "first order effects" of preference maximization are taken care of by the fact that f is strictly proper. In other words, there is no q,p with $q \neq p$ for which $f(q) \leq f(p)$. So, decisions about which p^* to announce are based on the second order effects of preference for dispersion. Roughly, the more the subject cares about the worst-case scenario (*i.e.* the closer they are to having a utility function $U(x) = \min_{\omega\in\Omega} x_\omega$), the flatter their corresponding indirect utility function. Hence, the more they will tend to minimize V in their behavior, and equate payoffs across states.

On the interpretation of p^*, taking the viewpoint of an experimenter who can conduct a single experiment because of complementarity issues, the announced p^* reflects a local notion of "belief" at $f(p^*)$ inasmuch as the subject is willing to face the act $f(p^*)$ given state-prices p^*. With a single data point, this is probably as good a notion as "belief" as any other.

IV. Examples

Examples follow. Most of the characterizations of the G functions here are taken directly from Cerreia-Vioglio et al. (2011b).

Example 1 (Translation Invariant, or Variational Preferences). Consider the variational preferences model of Maccheroni et al. (2006); applied to our setting, these preferences are those which can be written as preferences which are translation invariant in the sense that for all $x \in \mathbb{R}^\Omega$ and all $t \in \mathbb{R}$, $U(x + t(1,\ldots,1)) = U(x) + t$. In this case, it is easy to see that we can write $G(p,w) = w + c(p)$, for some proper convex (possibly infinite-valued) lower semicontinuous function c. This model incorporates the multiple priors model in the case where $c(p)$ is the convex-analytic indicator function of a convex set of priors P.[18] Writing down our formula, we want to solve

$$\arg\min_{p^*\in\Delta(\Omega)} \left(V(p^*)+c(p^*)\right).$$

Here, c has the interpretation of a *certainty equivalent*: $c(p)-c(q)$ measures the sure amount a decision maker holding only a riskless asset would pay to move from state prices q to prices p. To understand this claim, consider a decision maker maximizing $U(x)$ subject to $p\cdot x \le w$, where U has a certainty equivalent form (so that $U(x_1+t,\dots,x_n+t)=U(x)+t$). We observe that $c(p)=\max_{x:p\cdot x\le p\cdot 0} U(x)$. Hence, measured in terms of money, $c(p)$ is the value of facing prices p when endowment is 0 (riskless). To move from prices q to p, the individual would offer to pay $c(p)-c(q)$. And this willingness to pay is independent of riskless wealth, by the variational form.

For the special case of multiple priors, $c=0$ on the set of priors P, and is otherwise infinite. Hence, a multiple priors agent will always announce the probability in the set of priors which uniquely minimizes the value function $V(p^*)$; this is one of the main results of Grünwald and Dawid (2004). Thus, if the experimenter chooses a value function V which is strictly increasing along some direction in $\Delta(\Omega)$, then the agent will report the prior in their set that is highest in that direction. By varying the direction of increase, the experimenter can discover the entire set of priors. If the set of priors contains only the single prior p^* (the agent is an expected utility maximizer with belief p^*) then c is infinite-valued everywhere except at p^*. Hence, a risk-neutral subjective expected utility agent always announces their true belief p^*.

Example 2 (Subjective expected utility: CARA). For another special case, consider the subjective expected utility maximizer with CARA utility index $u(x)=-\exp(-\alpha x)$ where $\alpha>0$ and subjective probability π. We can consider the certainty equivalent utility representation of this preference:

$$U(x)=\alpha^{-1}\ln\left(\sum_{\omega\in\Omega} u(x_\omega)\pi(\omega)\right).$$

It turns out that the G function used in this environment has a very special form. First, it is a special case of the variational preference model (as CARA means that a preference is additively separable and translation invariant, the latter being the main characteristic of the variational model). Second, it is a special case for which the function $c(p)$ is given by a scaling of a *relative entropy* or *Kullback Leibler* function.

Formally, if q is absolutely continuous with respect to π, say that $R(q\parallel\pi)=\sum_{\omega:q(\omega)>0}\log\left(\frac{q(\omega)}{\pi(\omega)}\right)q(\omega)$, and otherwise $R(q\parallel\pi)=\infty$. CARA-preferences with parameter $\alpha>0$ and subjective probability π have a representation with the following G function: $G(w,p)=w+\alpha^{-1}R(p\parallel$

[18]In other words, it equals 0 for $p\in P$ and $+\infty$ otherwise.

π). Hence, a CARA individual will always solve the optimal scoring rule announcement problem by choosing to minimize the function $V(p^*) + \alpha^{-1} R(p^* \parallel \pi)$. Observe that $R(p^* \parallel \pi) = 0$ only when $p^* = \pi$, so that when $\alpha \to 0$, the problem tends to a risk-neutral agent with subjective probability π.

The observation that the relative entropy function leads to CARA-style preferences in this context seems to have been first made by Strzalecki (2011).

Example 3 (CARA and multiple priors). One can combine CARA models with multiple prior style models. Suppose that we consider a risk-averse decision maker with ambiguity-style preferences, so that the individual has a utility index $u(x) = -\exp(-\alpha x)$ of the CARA form, and a set of priors $P \subseteq \Delta(\Omega)$. It is straightforward to establish that such an individual can be represented with a (certainty equivalent dual) indirect utility function where $c(p) = \inf_{q \in P} \alpha^{-1} R(p \parallel q)$.

Example 4 (Gorman polar form and generalized translation invariance). Gorman (1961) provided necessary and sufficient conditions for utility functions to have Engel curves which are straight lines.[19] We focus on the case in which the Engel curves are parallel across different prices, since our consumption space is unbounded both above and below. Gorman originally proposed the family in order to meaningfully talk about "representative consumers," though they also have a natural interpretation of utility functions for which wealth effects are absent with respect to some "numeraire" bundle. That is, we want there to be a numeraire bundle $\beta \in \mathbb{R}^\Omega$ for which for all $t \in \mathbb{R}$ and all $x \in \mathbb{R}^\Omega$, $U(x + \beta t) = U(x) + t$.

Gorman called his utility functions *polar form*, since they are defined in terms of indirect utility. Therefore these form a natural class where the duality result is useful.

Formally, let us take $\beta \in \mathbb{R}^\Omega$, where $\beta \geq 0$ and $\beta \neq 0$. Let $\beta_+ = \{x \in \mathbb{R}^\Omega_+ : \sum_\omega x_\omega = 1\}$. Let $c : \beta_+ \to \mathbb{R} \cup \{+\infty\}$ be a proper, lower semicontinuous and convex function. Then define

$$G(p, w) = \begin{cases} \frac{w}{\beta \cdot p} + c\left(\frac{p}{\beta \cdot p}\right) \text{ if } \beta \cdot p \neq 0 \\ +\infty \text{ otherwise} \end{cases}.$$

This specifies an indirect utility in Gorman polar form. Note the obvious connection with the class of translation invariant preferences, described above. Translation invariant preferences are those which are in Gorman polar form with $\beta = (1, \ldots, 1)$.

V. Application: Bounding Misreports

Consider the CARA model described above in example 2 in a subjective expected utility framework. Nontrivial risk attitudes lead individuals to misreport their true subjective

[19] An *Engel curve* is the set of Walrasian demands for a fixed price as wealth varies.

belief when facing proper scoring rules. Our goal here is to construct a proper scoring rule in which misreports are bounded by some tolerance level, meaning they always fall within some acceptable distance of the true belief.

Specifically, let us define, for a given $\alpha > 0$ and probability measure π, the utility function:

$$U_\pi^\alpha(x) = -\sum_\omega \exp(-\alpha x_\omega)\pi(\{\omega\}).$$

For two measures $p, q \in \Delta(\Omega)$, we let $d(p,q) = \max_{E \subseteq \Omega} |p(E) - q(E)|$ represent the total variation metric. We use the metric d to measure the degree of misreporting.

Theorem 3. For every $\alpha^* \geq 0$, every $\epsilon > 0$, and every proper scoring rule f, there is $\gamma > 0$ for which for all $0 \leq \alpha \leq \alpha^*$ and all $\pi \in \Delta(\Omega)$, the solution:

$$p(\alpha, \pi, \gamma) = \arg\max U_\alpha^\pi(\gamma f(p))$$

satisfies $d(p(\alpha, \pi, \gamma), \pi) < \epsilon$.

The proof is in Appendix B. The interpretation of this result is that, so long as we are willing to assume risk aversion is bounded above, we can always guarantee that we have elicited subjects' true subjective probabilities within some prespecified tolerance. This is done simply by shrinking the stakes of the scoring rule to a sufficiently small level. Observe that the bound ϵ is uniform across all possible π.

A practical issue with Theorem 3 is that if ε is small then the scaling parameter γ may be very close to zero. This 'flattens' the incentives and may cause actual subjects to exhibit noisier behavior. We show next that if we rule out this rescaling procedure—in an effort to keep incentives strong—there is at least some non-empty range of risk aversion parameters for which optimal announcements will still be within ε of the truth. Here we let $p(\alpha, \pi) = p(\alpha, \pi, 1)$ represent the optimal announcement when the scoring rule is not scaled ($\gamma = 1$).

Corollary 2. For every $\epsilon > 0$ and every proper scoring rule f, there is some $\alpha^* > 0$ such that for all $0 \leq \alpha \leq \alpha^*$ and all $\pi \in \Delta(\Omega)$, the solution:

$$p(\alpha, \pi) = \arg\max U_\alpha^\pi(f(p))$$

satisfies $d(p(\alpha, \pi), \pi) < \epsilon$.

As mentioned in the Introduction, Kadane and Winkler (1988) and Armantier and Treich (2013) provide somewhat similar results, but only for the quadratic scoring rule. And our bound ε is a uniform bound, in the sense that it applies to all utility indices in the class and all possible beliefs.

VI. Conclusion and related literature

The literature on scoring rules is vast; the first characterization of proper scoring rules in terms of subdifferentials of convex functions was provided by McCarthy (1956); see also Savage (1971) and Fang et al. (2010). Gneiting and Raftery (2007) provides a survey of the literature up to 2007.

Theorem 1 has precedence in the literature. Probably the first such result is due to Winkler and Murphy (1970), who study the quadratic scoring rule for general expected utility preferences.[20] Kadane and Winkler (1988) calculate an explicit formula that expected utility maximizers would use when they have nontrivial risk attitudes and face a quadratic scoring rule. Grünwald and Dawid (2004) uncovers the special case of this result in the context of risk-neutral multiple priors (Chambers, 2008 later derives the same result). Offerman et al. (2009) derive a related result in a general decision-theoretic model (not necessarily quasiconcave preferences) essentially with a binary state space and the quadratic scoring rule, establishing that if a subject announces a probabilistic belief of an event which is not equal to 1/2, then we can say something about that optimal announcement.[21] Our result applies to any continuous strictly incentive compatible scoring rule; all of the classical ones as well as lesser known ones (for example Winkler, 1994). Furthermore, our result also applies to decision makers for which a meaningful concept of probability cannot even be defined.

Bickel (2007) establishes properties of individuals with CARA-style expected utility preferences: for example, he shows that one can add a constant payoff to each action in the profile of a scoring rule without changing behavior. He attributes this to what he calls the "delta" property; something economists would call translation invariance or quasilinearity, and is characterized by the variational model. This result follows from a straightforward application of Theorem 1. Jose et al. (2008) discusses a duality related to that of Grünwald and Dawid (2004), but with a different aim. There, they want to understand how a risk-averse expected utility maximizer will "bet" against a given set of priors.

There is a recent and elegant literature—pioneered by Maccheroni et al. (2006) and Cerreia-Vioglio et al. (2011b)—which exploits the duality between indirect and direct utility in order to study properties of uncertainty aversion. The framework is different; namely, they work with the richer structure of Anscombe and Aumann (1963) acts. The extra mathematical structure allows for the separate study of uncertainty and risk. The

[20] There are also elicitation mechanisms that are robust to risk attitudes (Grether, 1981, Karni, 2009, e.g.), though they do require probabilistic sophistication.

[21] The number 1/2 comes from the symmetry of the quadratic scoring rule.

duality investigated in Cerreia-Vioglio et al. (2011b) and discussed in detail in Cerreia-Vioglio et al. (2011a) is exactly the one we use here.[22] Though this work is concerned with uncertainty aversion, many of the mathematical results and characterizations appearing there apply verbatim here, and we have heavily borrowed from this work. Further, because of risk attitudes, it is often advocated that individuals be paid "in probabilities" of a good outcome, instead of monetary terms. This is based on the idea that, over purely risky prospects, individuals will likely conform to expected utility behavior (indeed this is the framework upon which the analysis of Anscombe and Aumann (1963) is built). The practical implementation of this idea in experiments is due to Roth and Malouf (1979). In such a framework, the framework of Cerreia-Vioglio et al. (2011b) is the appropriate one for studying elicitation questions. The natural counterparts of the following results hold as stated. We focus here on monetary payoffs for simplicity.

Similarly, one could restrict the domain of scoring rules to take only nonnegative values, and describe "homogeneous" utility indices U; these would correspond to G functions for which $G(p,w) = wG(p,1)$. It is likely that similar representations for CRRA preferences could derived.

Comparative statics on the G function in terms of the "more risk averse" relation of Yaari (1969) can be provided, and exist in Cerreia-Vioglio et al. (2011b) (in the form of comparative statics on uncertainty aversion).

This paper has added to the previous papers by considering a highly general specification of utility. Most previous works assume either that preference is expected utility, or consider the special case of "risk-neutral" multiple priors (Grünwald and Dawid (2004)). In contrast, our preferences never need to reference any concept of likelihood or state-contingent utility payoffs whatsoever. We leave the question of infinite state spaces, and elicitation mechanisms utilizing objective randomization devices to future research.

References

Aliprantis, C. D. and K. C. Border (1999). *Infinite Dimensional Analysis* (Second ed.). Springer.

Anscombe, F. J. and R. J. Aumann (1963). A definition of subjective probability. *Annals of mathematical statistics 34*(1), 199–205.

Armantier, O. and N. Treich (2013). Elicting beliefs: Proper scoring rules, incentives, stakes and hedging. *European Economic Review 62*, 17–40.

Azrieli, Y., C. P. Chambers, and P. J. Healy (2012). Incentives in experiments: a theoretical analysis. Working paper.

[22]Formally, they apply the duality to von Neumann-Morgenstern utils, and work with a much more general framework.

Berge, C. (1997). *Topological Spaces*. Dover.

Bickel, J. E. (2007). Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis 4*(2), 49–65.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011a). Complete monotone auasiconcave duality. *Mathematics of Operations Research 36*(2), 321–339.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011b). Uncertainty averse preferences. *Journal of Economic Theory 146*(4), 1275–1330.

Chambers, C. P. (2008). Proper scoring rules for general decision models. *Games and Economic Behavior 63*(1), 32–40.

Cornes, R. (1992). *Duality and Modern Economics*. Cambridge University Press.

Cover, T. M. and J. A. Thomas (2012). *Elements of Information Theory*. John Wiley & Sons.

de Finetti, B. (1949). Sulle stratificazioni convesse. *Annali di Matematica Pura e Applicata 30*, 173–183.

Fang, F., M. B. Stinchcombe, and A. B. Whinston (2010). Proper scoring rules with arbitrary value functions. *Journal of Mathematical Economics 46*(6), 1200–1210.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 107–114.

Gorman, W. M. (1961). On a class of preference fields. *Metroeconomica 13*(2), 53–56.

Grether, D. (1981). Financial incentive effects and individual decision making. Unpublished manuscript.

Grünwald, P. D. and A. P. Dawid (2004). Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics 32*(4), 1367–1433.

Jose, V. R. R., R. F. Nau, and R. L. Winkler (2008). Scoring rules, generalized entropy, and utility maximization. *Operations Research 56*(5), 1146–1157.

Kadane, J. B. and R. L. Winkler (1988). Separating probability elicitation from utilities. *Journal of the American Statistical Association 83*(402), 357–363.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica 77*(2), 603–606.

Konüs, A. A. (1939). The problem of the true index of the cost of living. *Econometrica 7*(1), 10–29.

Lau, L. J. (1969). Duality and the structure of utility functions. *Journal of Economic Theory 1*(4), 374–396.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica 74*(6), 1447–1498.

Machina, M. J. and D. Schmeidler (1992). A more robust definition of subjective probability. *Econometrica 60*(4), 745–780.

McCarthy, J. (1956). Measures of the value of information. *Proceedings of the National Academy of Sciences 42*(9), 654–655.

Nyarko, Y. and A. Schotter (2002). An experimental study of belief learning using elicited beliefs. *Econometrica 70*(3), 971–1005.

Offerman, T., J. Sonnemans, G. Van de Kuilen, and P. P. Wakker (2009). A truth serum for non-Bayesians: Correcting proper scoring rules for risk attitudes. *The Review of Economic Studies 76*(4), 1461–1489.

Roth, A. E. and M. W. Malouf (1979). Game-theoretic models and the role of information in bargaining. *Psychological Review 86*(6), 574.

Roy, R. (1947). La distribution du revenu entre les divers biens. *Econometrica 15*(3), 205–225.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association 66*(336), 783–801.

Shepherd, R. W. (1970). *Theory of Cost and Production Functions*. Princeton University Press.

Strzalecki, T. (2011). Axiomatic foundations of multiplier preferences. *Econometrica 79*(1), 47–73.

Ville, J. and P. K. Newman (1951-1952). The existence-conditions of a total utility function. *Review of Economic Studies 19*(2), 123–128.

Weymark, J. A. (1980). Duality results in demand theory. *European Economic Review 14*(3), 377–395.

Winkler, R. L. (1994). Evaluating probabilities: Asymmetric scoring rules. *Management Science 40*(11), 1395–1405.

Winkler, R. L. and A. H. Murphy (1970). Nonlinear utility and the probability score. *Journal of Applied Meteorology 9*(1), 143–148.

Yaari, M. E. (1969). Some remarks on measures of risk aversion and on their uses. *Journal of Economic theory 1*(3), 315–329.

Appendix A. Proof of Theorem 1

As a first step, consider the set $K = \mathrm{co}(f(\Delta(\Omega)))$ (the convex hull of the image of f). Observe that K is itself compact, since $\mathbb{R}^{\Omega}$ is finite-dimensional (Corollary 5.18 of Aliprantis and Border (1999)). We will show that there is a unique maximizer of f across the set K, and that this maximizer coincides with $\arg\max_p U(f(p))$.

So let $x^* \in \arg\max_{x \in K} U(x)$. Such a maximizer exists due to continuity of f and compactness of K. Let $Y = \{y : U(y) > U(x^*)\}$, which is open (by continuity) and convex (by quasiconcavity). The sets K and Y can therefore be separated by a hyperplane (Theorem 5.50 of Aliprantis and Border (1999)). Clearly, the hyperplane can be normalized to have direction in $\Delta(\Omega)$, by the fact that U is increasing. Let us call this direction p^*. Observe that the hyperplane p^* passes through x^*, as for any $\epsilon > 0$, $x^* + \epsilon(1,\ldots,1) \in Y$. Hence, conclude that for all $x \in K$, $p^* \cdot x \leq p^* \cdot x^*$; *i.e.* x^* maximizes $p^* \cdot x$ subject to $x \in K$. Clearly $f(p^*) \in K$ satisfies this inequality. We claim that it is the unique such element of K. So, let $\hat{x} \in K$, where $\hat{x} \neq f(p^*)$. Then there are $p_1,\ldots,p_n \in \Delta(\Omega)$, not all equal to p^* and $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$ for which $\hat{x} = \sum_i \lambda_i f(p_i)$. But by strict incentive compatibility, we then obtain $p^* \cdot \hat{x} < p^* \cdot f(p^*)$, contradicting the fact that $\hat{x}$ maximizes $p^* \cdot x$ subject to $x \in K$. So $x^* = f(p^*)$. Further, $f(p^*)$ is the unique maximizer of U in K. To see this, observe that by continuity and monotonicity, the closure of Y is $\{y : U(y) \geq U(f(p^*))\}$. Hence if $x' \in \arg\max_{x\in K} U(x)$, then $p^* \cdot x' \geq p^* \cdot f(p^*)$, which we have shown to be impossible.

Since p^* separates K and Y, and again by continuity, we have that $U(y) \geq U(f(p^*))$ implies that $p^* \cdot y \geq p^* \cdot f(p^*)$. We claim that this implies $U(f(p^*)) = G(p^*, p^* \cdot f(p^*))$. To see this, suppose by means of contradiction that there is y for which $p^* \cdot y \leq p^* \cdot f(p^*)$ and $U(y) > U(f(p^*))$. By continuity of U, we conclude that there is y^* for which $p^* \cdot y^* < p^* \cdot f(p^*)$ and $U(y^*) > U(f(p^*))$, contradicting the fact that $U(y^*) \geq U(f(p^*))$ implies $p^* \cdot y^* \geq p^* \cdot U(f(p^*))$. So, $U(f(p^*)) = G(p^*, p^* \cdot f(p^*))$.

Finally, let $\hat{p} \neq p^*$. Then $\hat{p} \cdot f(p^*) < \hat{p} \cdot f(\hat{p})$, by strict incentive compatibility. Therefore, there is $\epsilon > 0$ for which $\hat{p} \cdot (f(p^*) + \epsilon(1,\ldots,1)) < \hat{p} \cdot f(\hat{p})$, and since U is monotonic, we then conclude that $G(\hat{p}, \hat{p} \cdot f(\hat{p})) > U(f(p^*))$.

Conclude then that p^* uniquely solves $\min_p G(p, p \cdot f(p))$.

Appendix B. Proof of Theorem 3 and Corollary 2

Observe that we can without loss discuss maximizing any monotonic transformation of U^π_α rather than U^π_α itself. To this end, use the representation of CARA preferences described in the example. Let V be the value function associated with f, and let $R(p \parallel \pi)$ be the relative entropy function described above. By Theorem 1, it follows that we are searching for $\gamma > 0$ for which the solutions to $\arg\min_p \gamma V(p) + \alpha^{-1} R(p \parallel \pi)$ satisfy the requisite property.

Next, for any γ, if $\alpha' > \alpha$, then

$$R(p(\alpha,\pi,\gamma) \parallel \pi) \leq R(p(\alpha',\pi,\gamma) \parallel \pi)$$

simply by minimization.[23]

Similarly, if $\gamma' < \gamma$ we have $R(p(\alpha,\pi,\gamma) \parallel \pi) < R(p(\alpha,\pi,\gamma') \parallel \pi)$.

[23]Suppose by means of contradiction that

$$R(p(\alpha',\pi,\gamma) \parallel \pi) < R(p(\alpha,\pi,\gamma) \parallel \pi).$$

Next, observe that if $\gamma = 0$, π uniquely minimizes $\alpha^{-1}R(p \parallel \pi)$ across p, and further that $R(\pi \parallel \pi) = 0$.[24] Finally, by a Maximum Theorem-style argument, we know that $R(p(\alpha,\pi,\gamma) \parallel \pi)$ is continuous in π for each α, γ and that $R(p(\alpha,\pi,\gamma) \parallel \pi) \to 0$ as $\gamma \to 0$.[25] = Viewed as a function of π, we then observe that for any α, $R(p(\alpha,\pi,\gamma)) \to 0$ as $\gamma \to 0$; hence, we have a sequence of continuous functions which converge monotonically on a compact set; by Dini's Theorem (Berge (1997), p. 106) we may choose γ so that $R(p(\alpha^*,\pi,\gamma) \parallel \pi) < \frac{\epsilon^2}{2\ln(2)}$ for all π.

Now, observe that for all $\alpha \le \alpha^*$, we have

$$d(p(\alpha,\pi,\gamma),\pi) \le \sqrt{2\ln(2) R(p(\alpha,\pi,\gamma) \parallel \pi)} \le \epsilon,$$

where the first inequality follows by Pinsker's inequality and monotonicity in α, as described above (We have used the version of Pinsker's inequality from Cover and Thomas (2012), Lemma 11.6.1).

To see the proof of Corollary 2, let $\beta^* > 0$ be arbitrary, let f be a proper scoring rule, and let $\epsilon > 0$. Observe that by Theorem 3, there is $\gamma > 0$ such that for all $0 \le \beta \le \beta^*$, we have $d(p(\beta,\pi,\gamma),\pi) \le \epsilon$. Now, observe that for every $0 \le \beta \le \beta^*$, we have that the solution to the optimization problem:

$$\arg\max_{p \in \Delta(\Omega)} \sum_{\omega \in \Omega} -\exp(-\beta\gamma f(p)(\omega)),$$

namely $p(\beta,\pi,\gamma)$ coincides with $p(\beta\gamma,\pi)$. Set $\alpha^* = \beta^*\gamma$ and observe that the corollary follows.

Then we have

$$(\alpha^{-1} - \alpha'^{-1})R(p(\alpha',\pi,\gamma) \parallel \pi) < (\alpha^{-1} - \alpha'^{-1})R(p(\alpha',\pi,\gamma) \parallel \pi)$$

and

$$\gamma V(p) + \alpha'^{-1}R(p(\alpha',\pi,\gamma) \parallel \pi) \le \gamma V(p) + \alpha'^{-1}R(p(\alpha,\pi,\gamma) \parallel \pi),$$

so that

$$\gamma V(p) + \alpha^{-1}R(p(\alpha',\pi,\gamma) \parallel \pi) < \gamma V(p) + \alpha^{-1}R(p(\alpha,\pi,\gamma) \parallel \pi),$$

a contradiction.

[24] This does not mean that if there are no incentives, π will be the uniquely optimal choice for an individual.

[25] For the Maximum Theorem, see Berge (1997), p. 116. The result there does not apply verbatim since R can be infinite-valued. However, it is straightforward to establish that $p(\alpha,\pi,\gamma)$ is continuous in both π and γ. Thus, suppose that $(\pi_n,\gamma_n) \to (\pi^*,\gamma^*)$. Take any subsequence $p(\alpha,\pi_{n_k},\gamma_{n_k})$, and let $p(\alpha,\pi_{n_{k_j}},\gamma_{n_{k_j}})$ be a convergent subsequence, say to p^*. We will show that $p^* = p(\alpha,\pi^*,\gamma^*)$. So, let $\hat{p}$ be arbitrary; observe that

$$\gamma_{n_{k_j}} V(p(\alpha,\pi_{n_{k_j}},\gamma_{n_{k_j}})) + \alpha^{-1}R(p(\alpha,\pi_{n_{k_j}},\gamma_{n_{k_j}}) \parallel \pi_{n_{k_j}}) \le \gamma_{n_{k_j}} V(\hat{p}) + \alpha^{-1}R(\hat{p} \parallel \pi_{n_{k_j}})$$

and take limits, using the fact that V is continuous and R continuous on its effective domain, to establish that $p(\alpha,\pi_{n_{k_j}},\gamma_{n_{k_j}}) \to p(\alpha,\pi^*,\gamma^*)$ Since every subsequence of $p(\alpha,\pi_n,\gamma_n)$ has a sub-subsequence which converges to $p(\alpha,\pi^*,\gamma^*)$, we establish that $p(\alpha,\pi_n,\gamma_n) \to p(\alpha,\pi^*,\gamma^*)$.