Axiomatic Foundations of Agency Theory

Edi Karni*

Department of Economics

Johns Hopkins University

Baltimore, MD 21209, USA

July 27, 2004

Abstract

This paper develops axiomatic foundations for the parameterized distribution formulation of agency theory. Both the principal and the agent are subjective expected utility maximizers. Unlike in traditional subjective expected utility theory, the subjective probabilities are defined directly on the outcomes and are action dependent.

JEL Calssification: D81, D82

Keywords: Agency theory, Contracts theory, Moral Hazard

*I am grateful to Simon Grant, Bernard Salanié, two anonymous referees and, in particular, to Peter Wakker for their comments and suggestions, and to the NSF for financial support under grant SES-0314249.

1 Introduction

The ability of decision makers to influence, by their (unobserved) actions, the likely outcomes of risky enterprises is an essential ingredient of the principal-agent relationship in the presence of moral hazard. The analysis of this relationship often invokes the parameterized distribution formulation, pioneered and popularized by Mirrlees (1974, 1976).[1] Despite its widespread use, the axiomatic underpinnings of the conducts of the principal and the agent in the context of the parameterized distribution formulation have not yet been explored. This paper fills this gap. As usual, the study of the axiomatic foundations is intended to uncover the assumptions implicit in the specification of the objective functions of the parties involved. The understanding of these assumptions helps evaluate the plausibility of the model and provides insight to its behavioral underpinning.[2]

In the parameterized distribution formulation of agency theory, at the primitive level, there is a set of outcomes and a family of distribution functions on these outcomes, whose members are parameterized by the agent's action. Thus in contrast to the state space formulation, the parameterized distribution formulations of the principal-agent relations dispense with use of a state space as part of the analytical framework. To remain faithful to the spirit of this approach, the axiomatization of agency theory in this paper avoids the use of a state

[1] See Hart and Holmstrom (1979), Chambers and Quiggin (2000).

[2] This is analogous to the axiomatization of Bernoulli's expected utility hypothesis. In addition to clarifying the principles of choice behavior underlying Bernoulli's hypothesis, thereby allowing the assessment of their reasonableness, the study of the axiomatic structure suggested direct tests of the theory.

space.

A principal-agent relationship is governed by a contract specifying the agent's monetary payoff contingent on the outcomes. Since the agent's choice of action affects the likely realization of the alternative outcomes, the contracts are designed, by the principal, to induce the agent to implement an action that is in the principal's best interest. Naturally the roles of the two parties in the relationship are asymmetric. Typically the principal's concerns include both the outcome (e.g., output) and the agent's monetary compensation, while the agent is concerned with the outcome only to the extent that it affects his payoff. Moreover, the choice of action affects the agent's well-being directly, (e.g., fatigue caused by effort) while it affects the well-being of the principal only indirectly through is influence on the likely realization of different outcomes. Consequently, at the axiomatic level, the behaviors of the principal and the agent call for distinct treatments.

At the primitive level, the principal-agent problem may be stated as follows: Let $\mathcal{A}$ be a feasible set of actions, and denote by X the (arbitrary) set of outcomes. A contract is a real-valued function, w, on X specifying the agent's monetary reward as a function of the observed outcome. Denote by W the set of contracts. The principal and the agent have preference relations on the set $\mathcal{A} \times W$ of action-contracts pairs that are denoted by $\succcurlyeq^P$ and $\succcurlyeq^A$, respectively. The principal's problem may be stated as follows: Choose an action-contract pair (a', w') such that

$$(a', w') \succcurlyeq^P (a, w) \text{ for all } (a, w) \in \mathcal{A} \times W \tag{1}$$

subject to the incentive compatibility constraints

$$(a', w') \succsim^A (a, w') \text{ for all } a \in \mathcal{A}, \tag{2}$$

and the participation constraint

$$(a', w') \succsim^A (\bar{a}, \bar{w}), \tag{3}$$

where $(\bar{a}, \bar{w})$ denotes the agent's "outside option," (that is, the action-contract pair that is equivalent to the best course of action available to the agent if he rejects the contract).

If the principal and the agent are subjective expected utility maximizers and, in addition, the agent's utility is additively separable in money and actions then the principal-agent problem may be reformulated as follows. Let u^P denote the principal's utility function and, for each $a \in \mathcal{A}$ let $\pi^P(\cdot; a)$ denote the principal's subjective probabilities on X. Similarly, let u^A and v denote the agent's utility functions of money and actions, respectively, and let $\pi^A(\cdot; a)$ be his action-dependent subjective probabilities on X. For the purpose of mathematical analysis, the principal-agent problem may be stated as follows: Choose an action-contract pair (a', w'),

$$(a', w') \in \arg\max_{\mathcal{A} \times W} \sum_{x \in X} u^P(w(x); x)\, \pi^P(x; a) \tag{4}$$

subject to the incentive compatibility constraints, for all $a \in \mathcal{A}$,

$$\sum_{x \in X} u^A(w'(x))\, \pi^A(x; a') + v(a') \geq \sum_{x \in X} u^A(w'(x))\, \pi^A(x; a) + v(a) \tag{5}$$

and the participation constraint

$$\sum_{x \in X} u^A(w'(x))\, \pi^A(x; a') + v(a') \geq \bar{u}, \tag{6}$$

where $\bar{u}$ denotes the utility of the "outside option."

With these observations in mind the question addressed in this paper may be stated as follows: What must be true about the preference relations of the principal and the agent for the programs (1)–(3) and (4)-(6) to be equivalent? The difficulty posed by this question is due, in part, to the fact that dispensing with the state-space constitutes a major departure from the accepted theories of decision making under uncertainty and requires a new analytical framework. Put differently, the representations in (4)-(6) are neither implied by the existing subjective expected utility theory nor do they constitute an obvious extension of it. Note, in particular, that since the application of subjective expected utility theory (e.g., Savage [1954] or Anscombe and Aumann [1963]) would require that the utility functions be outcome independent and the subjective probabilities be action independent, simply replacing states with outcomes will not do. (The principal's utility-of-money function in (4) is outcome-dependent, and both the principal's and the agent's objective functions involve action-dependent subjective probabilities. These characterizations are incompatible with the exigencies of subjective expected utility theory).

A different issue concerns the meaning of the common beliefs assumption that is often used in the analysis of principal-agent relationships. Specifically, the agent's subjective expected utility representation is based on the axiom of outcome-independent preferences. This axiom does not entail the outcome-independent utility functions that figure in the representation in (4)-(6). In other words, the utility functions may be outcome-dependent provided they are linear transformations of one another. The imposition of outcome-independent utility functions lacks behavioral justification and might confound beliefs and valuations in the representation of the agent's preferences and, as a result, renders the interpretation of the common beliefs assumption ambiguous.

In this paper I invoke the analytical framework developed in Karni (2004a) and, within this framework, I identify necessary and sufficient conditions for the existence of representations of the principal's and the agent's preferences entailing unique families of action-dependent subjective probabilities quantifying the their beliefs regarding the likely realization of outcomes conditional on the agent's choice of action. Moreover, I provide a choice-theoretic method of verifying that the agent's utility functions are indeed outcome independent and, thereby, eliminate the ambiguity regarding the interpretation of the common beliefs assumption.

The issue of axiomatizing subjective expected utility theory with state-dependent preferences and moral hazard was first addressed by Drèze (1961), (1987). The work of Drèze differs fundamentally from that of this paper and is discussed in more details in section 5, where it could more readily be contrasted with the present analytical framework and results.

The analytical framework is introduced in the next section. In sections 3 and 4, respectively, I develop a choice-theoretic axiomatic models of the principal's and the agent's behavior. The model of the principal's behavior is based on the decision theory developed in Karni (2004a). The model of the agent's behavior is novel. Section 5 contains a discussion of related literature. The proofs appear in Section 6.

2 The Analytical Framework

2.1 Preliminaries

Let X be a finite set of *outcomes,* and let $\mathcal{A}$ be a topological space whose elements represent *actions* that may be chosen by the agent.[3] A *contract*, w, is a mapping from X to $\mathbb{R}$, the set of real numbers. A contract specifies a monetary transfer from the principal to the agent, contingent on the realized outcome. Let W denote the set of all contracts (that is, $W = \mathbb{R}^X$) and assume that it is endowed with the $\mathbb{R}^{|X|}$ topology. Denote by (w_{-x}, r) the contract obtained from $w \in W$ by replacing the x–coordinate of w, that is, $w(x)$, with r. Similarly, for each $Y \subset X$ and $w, w' \in W$, let $w_Y w'$ be the contract in W defined by $(w_Y w')(x) = w(x)$ for all $x \in Y$ and $(w_Y w')(x) = w'(x)$ for all $x \in X - Y$. Two contracts, say w and w', are said to *agree* on $Y \subset X$ if $w(x) = w'(x)$ for all $x \in Y$.

Examples: (1) Health insurance - Outcomes represent the states of a person's health, actions correspond to alternative exercise and diet regimes, and contracts represent alternative health insurance policies. (2) Sharecropping - Outcomes are alternative levels of output, actions correspond to time and effort spent working, and the contracts are sharecropping agreements.

[3] In the state-space formulation of agency theory a technology, represented by a mapping $T : \mathcal{A} \times S \rightarrow X$, assigns a unique outcome to each action-state pair. For an axiomatization of the the state-space formulation of agnecy theory see Karni (2004).

Decision makers, in their roles of principals or agents, are suppose to be able to evaluate alternative action-contracts pairs. They are aware that a choice of an action, a, may affect the likely realization of different outcome in X. Consequently the *choice set,* $\mathbb{C}$, consists of all the action-contract pairs (that is, $\mathbb{C} := \mathcal{A} \times W$) and is assumed, henceforth, to be endowed with the product topology. A choice of an action a and a contract w ultimately results in an outcome-payoff pair, $(x, w(x))$. I refer to outcome-payoff pairs as *consequences* and denote by C the set of all consequences (that is, $C := X \times \mathbb{R}$).

The principal and the agent are characterized by the preference relations, $\succcurlyeq^P$ and $\succcurlyeq^A$ on $\mathbb{C}$. Given a preference relation $\succcurlyeq^h, h \in \{P, A\}$, the strict preference relation, $\succ^h$, and the indifference relation, $\sim^h$, are the asymmetric and symmetric parts of $\succcurlyeq^h$, respectively. For each $a \in \mathcal{A}$, the preference relation $\succcurlyeq^h$ on $\mathbb{C}$ induces a conditional preference relation on W, defined as follows: For all $w, w' \in W$, $w \succcurlyeq_a^h w'$ if and only if $(a, w) \succcurlyeq^h (a, w')$.

A decision maker may believe that if a particular action is selected, certain outcomes cannot occur. This belief manifests itself in indifference among all the contracts that agree on the set of all other outcomes. An outcome x is said to be *null given the preference relation* $\succcurlyeq^h$ *and the action* a if $(w_{-x}, r) \sim_a^h (w_{-x}, r')$ for all $r, r' \in \mathbb{R}$, otherwise x is *nonnull given preference relation* $\succcurlyeq^h$ *and the action* a. Denote by $X(a; \succcurlyeq^h)$ the subset of X that are nonnull given $\succcurlyeq^h$ and a.

It is customary in agency theory to suppose that all outcomes are possible under every feasible action.[4] This assumption is maintained throughout. Formally,

Full support assumption: For all $a \in \mathcal{A}$ and $h \in \{P, A\}$, $X\left(a; \succcurlyeq^h\right) = X$ and $\mid X \mid \geq 2$.

2.2 Common features of the preference relations

In general, the principal's and the agent's preferences are qualitatively distinct. Typically, the agent's well-being is directly affected by the action and is not directly affected by the outcome. In contrast, the principal's well-being is directly affected by the outcomes but not by the actions. In some respects, however, the preference relations of the principal and the agent are the same. To begin with, they are continuous weak orders. This is captured by the following two standard axioms that hold for $h \in \{P, A\}$.

(A.1) (**Weak order**) $\succcurlyeq^h$ *is a complete and transitive binary relation.*

(A.2) (**Continuity**) *For all* $(a, w) \in \mathbb{C}$ *the sets* $\{(a', w') \in \mathbb{C} \mid (a', w') \succcurlyeq^h (a, w)\}$ *and* $\{(a', w') \in \mathbb{C} \mid (a, w) \succcurlyeq^h (a', w')\}$ *are closed*

The next axiom, also assumed to hold for $h \in \{P, A\}$, requires that the "intensity of preferences" for monetary payoffs contingent on any given outcome be independent of the action that resulted in that outcome and, given an action, this intensity of preferences be

[4] If an outcome, say x, is null given the action, a, that the principal wants the agent to adopt, the principal could impose a large penalty on the agent if x obtains. This will deter the latter from taking actions under which x is nonnull (see Salanié [1997]).

independent of the payoffs associated with all other outcomes. It invokes Wakker's (1987) idea of cardinal consistency and, in its present form, is an adaptation of Karni's (2004a) action-independent betting preferences.

(A.3) (**Action independence**) *For all* $a, a' \in \mathcal{A}$, $w, w', w'', w''' \in W$, $x \in X$, *and* $r, r', r'', r''' \in \mathbb{R}$, *if* $(a,(w_{-x},r)) \succcurlyeq^h (a,(w'_{-x},r'))$, $(a,(w'_{-x},r'')) \succcurlyeq^h (a,(w_{-x},r'''))$, *and* $(a',(w''_{-x},r')) \succcurlyeq^h (a',(w'''_{-x},r))$ *then* $(a',(w''_{-x},r'')) \succcurlyeq^h (a',(w'''_{-x},r'''))$.

To grasp the meaning of outcome independence, think of the preferences $(a,(w_{-x},r)) \succcurlyeq^h (a,(w'_{-x},r'))$ and $(a,(w'_{-x},r'')) \succcurlyeq^h (a,(w_{-x},r'''))$ as indicating that, given action a and outcome x, the intensity of the preferences of r'' over r''' is sufficiently greater than that of r' over r as to reverse the preference ordering of the contingent payoffs w_{-x} and w'_{-x}. The axiom requires that these intensities not be contradicted when the action a' is replaced by a.

To grasp the role that action independence plays in the analysis that follows consider first two special cases. Case 1: $a = a'$. Axiom (A.3) is reduced to the generalized triple cancellation condition (see Wakker [1989] p. 70). Under the full support assumption, when the number of outcomes is at least three, action independence together with weak order and continuity constitute necessary and sufficient conditions for the existence of additive representation of $\succcurlyeq^h_a$. Case 2: $a = a', w' = w'', r' = r''$ and $\sim^h$ instead of $\succcurlyeq^h$in (A.3). Axiom (A.3) is reduced to the hexagon condition which, together with weak order and continuity constitute necessary and sufficient conditions for the existence of additive representation of

$\succcurlyeq_a^h$ when there are exactly two outcomes. Furthermore, as the next Lemma shows, action independence together with weak order and continuity constitute necessary and sufficient conditions for the existence of additive representations of $\succcurlyeq_a^h, a \in \mathcal{A}$, and, in addition, for these representations to be linear transformations of one another.

2.3 Additive representation

An array of real-valued functions $(v_s)_{s\in S}$ is said to be a *jointly cardinal additive representation* of a binary relation $\succeq$ on a product set $D = \Pi_{s\in S} D_s$ if, for all $d, d' \in D$, $d \succeq d'$ if and only if $\sum_{s\in S} v_s(d_s) \geq \sum_{s\in S} v_s(d'_s)$, and the class of all functions that constitute an additive representation of $\succeq$ consists of those arrays of functions, $(\hat{v}_s)_{s\in S}$, for which $\hat{v}_s = \lambda v_s + \zeta_s$, $\lambda > 0$ for all $s \in S$. The representation is continuous if the functions $v_s, s \in S$ are continuous.

The next lemma states necessary and sufficient conditions for the existence of jointly cardinal continuous additive representation of $\succcurlyeq_a^h$ on W, $a \in A$. Moreover, the additive value functions conditional on distinct actions are positive affine transformations of one another.

Lemma 1 *Suppose that the full-support assumption holds. Then the following conditions are equivalent:*

(i) The relation $\succcurlyeq^h$ on $\mathbb{C}$ satisfies (A.1)–(A.3).

(ii) For each $a \in \mathcal{A}$ there exist jointly cardinal continuous additive-value functions, $\{V_a(\cdot; x)\}_{x\in X}$

representing $\succcurlyeq_a^h$ *on* W*, such that, for all* $a \in \mathcal{A}$ *and* $x \in X$*,* $V_a(\cdot;x) = \lambda(x;a)V(\cdot;x) + \xi(x;a)$*, where* $\lambda(x;a) > 0$ *and* $V(\cdot;x)$ *is a real-valued function.*

3 Representations of the Principal's Preferences

3.1 Subjective expected utility representation

Constant-valuation bets describe outcome-contingent payoffs that, once fixed, make the decision maker indifferent among all the actions in $\mathcal{A}$ (see Karni [2004a]). In this paper bets are contracts and constant-valuation bets are constant-valuation contract. An example of such contract is an insurance policy providing coverage that renders the insured indifferent among all possible outcomes (that is, full insurance policy). Since the principal's ultimate concern are the consequences, implicit in the notion of constant-valuation contracts is the presumption that, insofar as the principal's well-being is concerned, given a constant-valuation contract, all the consequences are equivalent. That is, if w^* is a constant-valuation contract then, for all $x, x' \in X$, the $(x, w^*(x))$ and $(x', w^*(x'))$ are equivalent. For example, if the outcomes represent levels of revenues (that is, if $X \subset \mathbb{R}$) and the principal's sole concern is the level of profits, $x - w(x)$, then $\{w^* \mid w^*(x;c) = x - c, c \in \mathbb{R}\}$, is the set of constant-valuation contracts. For constant-valuation contracts to be well-defined requires richness of the set of actions. In particular, given $a \in A$ let $\mathcal{P}(a) = \{I(a;w) \mid w \in W\}$, where $I(a;w) = \{w' \in W \mid (a,w') \sim^P (a,w)\}$. Thus $\mathcal{P}(a)$ is a partition of W into equivalence classes given by the indifference relation $\sim^P$. Suppose that there are $w \in W$ for which $\cap_{a\in A} I(a;w)$ is a singleton set. In other words, there is a unique $w^* \in I(w;a)$ that I shall refer to as *constant-valuation contract* such that $(a,w^*) \sim^P (a',w^*)$, for all $a,a' \in \mathcal{A}$. I assume that their are two constant-valuation contracts, w^{**} and w^* such that $(a,w^{**}) \succ^P (a,w^*)$. By definition and transitivity, $(a,w^{**}) \succ^P (a,w^*)$ if and only if $(a',w^{**}) \succ^P (a',w^*)$, for all $a' \in \mathcal{A}$. Hence, without risk of ambiguity, I shall write $w^{**} \succ^P w^*$.

The next theorem asserts that the principal's preferences are representable by a subjective expected utility functional, with action-independent utilities and action dependent probabilities.

Theorem 2 *If there are at least two outcomes, the full-support assumption holds, and there exist two constant-valuation contracts, w^{**},$w^* \in W$ such that $w^{**} \succ^P w^*$, then*

(a) The following conditions are equivalent:

(a.i) The preference relation $\succcurlyeq^P$ on $\mathbb{C}$ satisfies (A.1)–(A.3).

(a.ii) There exist continuous function $u^P : C \to \mathbb{R}$ and a family of probability measures $\{\pi^P(\cdot;a)\}_{a\in\mathcal{A}}$ on X such that for all $(a,w),(a',w') \in \mathbb{C}$,

$$(a,w) \succcurlyeq^P (a',w') \Leftrightarrow \sum_{x\in X} u^P(w(x);x)\,\pi^P(x;a) \geq \sum_{x\in X} u^P(w'(x);x)\,\pi^P(x;a').$$

(b) The utility function u^P is unique up to a positive linear transformation.

(c) For each $a \in \mathcal{A}$, $\pi^P(\cdot;a)$ is unique and has full support.

Theorem 2 is a variation of Theorem 1 in Karni (2004a).

4 Representations of the Agent's Preferences

4.1 Outcome independence and conditional expected utility representations.

Agents preference relations over monetary payoff are assumed to be "outcome independent." The following axiom, which is similar to Wakker's (1987) cardinal consistency, captures this idea:

(A.4) (**Outcome independence**) *For all* $a \in \mathcal{A}$, $w, w', w'', w''' \in W$, $x, y \in X$, *and* $r, r', r''r''' \in \mathbb{R}$, *if* $\left(a, (w'_{-x}, r')\right) \succcurlyeq^A (a, (w_{-x}, r))$, $(a, (w_{-x}, r'')) \succcurlyeq^A \left(a, (w'_{-x}, r''')\right)$, *and* $\left(a, (w''_{-y}, r)\right) \succcurlyeq^A \left(a, (w'''_{-y}, r')\right)$, *then* $\left(a, (w''_{-y}, r'')\right) \succcurlyeq^A \left(a, (w'''_{-y}, r''')\right)$.

Outcome independence is analogous to action independence. In particular, the preferences $\left(a, (w'_{-x}, r')\right) \succcurlyeq^A (a, (w_{-x}, r))$ and $(a, (w_{-x}, r'')) \succcurlyeq^A \left(a, (w'_{-x}, r''')\right))$ indicate that, given the action a and the outcome x, the "intensity" of the preference for r'' over r''' is sufficiently greater than that of r' over r as to reverse the order of preference between the payoffs represented by the contracts w'_{-x} and w_{-x}. Outcome independence requires that these intensities not be contradicted by the preferences between the same payoffs given any other outcome y.

Lemma 3 below indicates the role of axiom (A.4) in the proofs of the agent's representation theorems, namely, lending the utility representations additively separable functional form.

Lemma 3 *Suppose the full-support assumption holds. Then the preference relation* $\succcurlyeq_a^A$ *on* W *is a continuous weak order satisfying (A.4) if and only if there exist continuous real-valued function,* $U(\cdot;a)$, *on* $\mathbb{R}$ *a probability measure* $p^A(\cdot;a)$ *on* X *such that, for all* $w, w' \in W$, *and* $a \in A$,

$$(a,w) \succcurlyeq^A (a,w') \Leftrightarrow \sum_{x \in X} U(w(x),a)\, p^A(x;a) \geq \sum_{x \in X} U(w'(x),a)\, p^A(x;a).$$

Moreover, $U(\cdot,a)$ *is unique up to a positive linear transformation and, given* $U(\cdot,a)$, *the probability measure* $p^A(\cdot;a)$ *is unique.*

Lemma 3 is implied by Theorem IV.2.7. in Wakker (1989). Note that the utility functions in Lemma 3 are outcome-independent.

4.2 The agent's subjective expected utility representation I

Let $F_w(r;a) = \sum_{y \in \{x \in X | w(x) \leq r\}} p^A(y;a)$. Then the contract w is said to dominate another contract w' according to first order stochastic dominance if $F_w(r;a) \leq F_{w'}(r;a)$ for all $r \in \mathbb{R}$. If, in addition, the inequality is strict for some r then w is said to strictly first-order dominate w'. I denote the first-order stochastic dominance relation by $w \geq_1 w'$ and its strict version by $w >_1 w'$. The next axiom requires that $\succcurlyeq_a^A$ be monotonic with respect to first order stochastic dominance.

(A.5) (**Monotonicity**) The preference relation $\succcurlyeq_a^A$ on W satisfies $w \succcurlyeq w'$ whenever $w \geq_1 w'$ and $w \succ w'$ whenever $w >_1 w'$.

The next theorem asserts that the agent's preference relation is representable by a subjective expected utility functional with outcome-independent, action-affine, utilities and action-dependent probabilities.

Theorem 4 *If the full-support assumption holds then:*

(a) The following conditions are equivalent:

(a.i) The preference relation $\succcurlyeq^A$ on $\mathbb{C}$ satisfies (A.1)–(A.5).

(a.ii) There exists a continuous, monotonic increasing, function $u^A : \mathbb{R} \to \mathbb{R}$, a function $v : \mathcal{A} \to \mathbb{R}$, a positive function $b : \mathcal{A} \to \mathbb{R}$, and a family of probability measures $\{p^A(\cdot; a)\}_{a \in \mathcal{A}}$ on X such that for all $(a, w), (a', w') \in \mathbb{C}$, $(a, w) \succcurlyeq^A (a', w')$ if and only if

$$\sum_{x \in X} b(a) u^A(w(x)) p^A(x; a) + v(a) \geq \sum_{x \in X} b(a') u^A(w'(x)) p^A(x; a') + v(a').$$

(b) If $\bar{b}, \bar{u}^A$, and $\bar{v}$ represent the preference relation $\succcurlyeq^A$ in the sense of (a.ii), then $\bar{b} = \gamma b + \zeta_b, \bar{u}^A = \gamma \bar{u}^A + \zeta_u(a)$, and $\bar{v} = \gamma v + \zeta$, where $\gamma > 0$, and, for all $a, a' \in \mathcal{A}$, $\zeta_u(a) b(a) = \zeta_u(a') b(a')$ and $\zeta_b(a) \sum_{x \in X} u^A(w(x)) \pi^A(x; a) = \zeta_b(a') \sum_{x \in X} u^A(w(x)) \pi^A(x; a')$.

(c) For each $a \in \mathcal{A}$, given $u^A(\cdot)$, $p^A(\cdot; a)$ is unique and has full support.

In contrast to the probabilities that figure in the representation of the principal's preferences (Theorem 2) the probabilities in Theorem 4 might confound the agent's beliefs and his valuation of the monetary payoffs associated with the contract. To grasp this assertion observe that the choice of an outcome-independent utility function is a convention that is not implied by the axioms.[5] For example, let η and ξ be real-valued functions on X, where η is a positive function. Suppose that the agent is a subjective expected utility maximizer whose beliefs are represented by the probability measures $\{\hat{p}^A(\cdot;a)\}_{a\in\mathcal{A}}$, his tastes are captured by a outcome-dependent utility function $\hat{u}^A(r,x) = u^A(r)\,\eta(x) + \xi(x)$, $(r,x) \in \mathbb{R}\times X$, and his (dis)utility of action is quantified by the real-valued function $\hat{v}$ on $\mathcal{A}$. Suppose that the agent's preferences are represented by:

$$(a,w) \longmapsto \sum_{x\in X} \left[u^A(w(x))\,\eta(x) + \xi(x)\right] \hat{p}^A(x;a) + \hat{v}(a)\,. \tag{7}$$

For each $a \in \mathcal{A}$ define $p^A(x;a) := \hat{p}^A(x;a)\,\eta(x)/b(a)$ for all $x \in X$, where $b(a) := \sum_{x\in X} \hat{p}^A(x;a)\,\eta(x)$. Let $v(a) = \sum_{x\in X} \xi(x)\,\hat{p}^A(x;a) + \hat{v}(a)$. Then the representation in (7) is reduced to that in Theorem 4, that is

$$(a,w) \longmapsto \sum_{x\in X} b(a)\,u^A(w(x))\,p^A(x;a) + v(a)\,. \tag{8}$$

Thus the utilities and probabilities ascribed to the agent do not necessarily represent his beliefs and tastes. In fact, an agent characterized by $\{\hat{u}^A, (\hat{p}^A(\cdot;a))_{a\in\mathcal{A}}, \hat{v})\}$ is behaviorally indistinguishable from agent characterized by $\{b, u^A, (p^A(\cdot;a))_{a\in\mathcal{A}}, v)\}$. This example also

[5] Drèze (1987); Schervish, Seidenfeldt, and Kadane (1990); Karni (1993), (1996), (2003); and Karni and Schmeidler (1993) make the same point in the traditional context of subjective expected utility theory.

shows that the coefficient $b(a)$ that appear in the representation, which is sometimes interpreted as a component of the utility if action (see, for example, Grossman and Hart [1983]) is in fact, at least partly, an artifact of the normalization.

It is customary in agency theory to assume that the principal and the agent agree on the parameterized probabilities. The discussion in this paper reveals that the meaning of this assumption is ambiguous. Specifically, if this assumption refers to *common beliefs,* then $\pi^P(\cdot \mid a) = \hat{p}^A(\cdot \mid a)$, where $\hat{p}^A(\cdot \mid a)$ is as equation (7) and if it refers to *common ascribed beliefs,* then $\pi^P(\cdot \mid a) = p^A(\cdot \mid a)$, where $p^A(\cdot \mid a)$ is as equation (8). The two interpretations of the common prior assumption give rise to two distinct formulations of the principal-agent problem.

Broadly speaking, the analysis of incentive contracts is based on the assumption that the agent's utility function is effect-independent. In other words, implicit in the usual formulations of the moral-hazard problem is the assumption that, insofar as the agent is concerned, constant-valuation contracts are in fact constant-payoff contracts (that is, if w^* is a constant-valuation contract then $w^*(x) = w^*(x')$ for all $x, x' \in X$). This observation raises the question: Is there a choice-theoretic meaning to the imposition of state-independent utility function? The answer to this question is the subject matter of the next subsection.

4.3 The agent's subjective expected utility representation II

In many situations involving principal-agent relationships it makes sense to suppose that the agent's utility function is outcome independent. However, as our discussion at the end of the preceding subsection suggests, this supposition seems to lack choice-theoretic meaning. The difficulty in trying to separate the agent's utilities and beliefs is that, in addition to its effect on the probability distribution on the outcomes, the choice of an action has direct effect on the agent's well-being. Therefore, the definition of constant valuation contracts that was used to disentangle beliefs from valuations in the case of the principal's, does not work in the case of the agent. A different approach is required.

To lend the hypothesis that the agent's utility function is outcome independent a choice-theoretic meaning I propose the following idea. If constant-payoffs contracts are constant-valaution contracts and the utility function is additively separable in money and action, then the application of Theorem 4 to the subset of constant-payoff contracts reveals that the expected utility part of the representation is independent of a, (that is, if $w(x) = r$ for all x then $\sum_{x \in X} u^A(w(x)) p^A(x; a) = u(r)$ regardless of a). This, in turn, implies that the utility function is outcome independent.

To formalize this idea I need to strengthen the topological assumption on $\mathcal{A}$. In particular, I assume that the set $\mathcal{A}$ is a connected separable topological space. (For example, elements of $\mathcal{A}$ may signify levels of effort in which case $\mathcal{A}$ may be represented by an interval in the real line.) Let $W^c \subset W$ be the subset of constant-payoff contracts (that is, constant functions) in

W. Thus W^c corresponds to $\mathbb{R}$. Then the separability of the utilities of actions and contracts is attained by the imposition of the hexagon condition restricted to the set $\mathcal{A} \times W^c$.

(A.6) (**Restricted hexagon condition**) *For all* $a, a', a'' \in \mathcal{A}$ *and* $w, w', w'' \in W^c$ *if* $(a, w') \sim^A (a', w)$ *and* $(a, w'') \sim^A (a', w') \sim^A (a'', w)$ *then* $(a', w'') \sim^A (a'', w')$.

A preference relation $\succsim^A$ on $\mathcal{A} \times W^c$ is said to satisfy *coordinate independence* if for all $a, a' \in \mathcal{A}$ and $w, w' \in W^c$, $(a, w) \succsim^A (a, w')$ if and only if $(a', w) \succsim^A (a', w')$ and $(a, w) \succsim^A (a', w)$ if and only if $(a, w') \succsim^A (a', w')$.

(A.7) (**Coordinate independence**) *The preference relation* $\succsim^A$ *on* $\mathcal{A} \times W^c$ *satisfies coordinate independence.*

The sets $\mathcal{A}$ and W^c are said to be *essential with respect to a preference relation* $\succsim^A$ if there are $(a, w), (a', w), (a, w), (a, w') \in \mathcal{A} \times W^c$ such that $(a, w) \succ^A (a', w)$ and $(a, w) \succ^A (a, w')$.

(A.8) (**Essentiality**) *The sets* $\mathcal{A}$ and W^c *are essential with respect to a preference relation* $\succsim^A$.

The following theorem is the agent's subjective expected utility representation analogous to Theorem 2

Theorem 5 *If the full-support assumption holds,* $\mathcal{A}$ *is a connected separable topological space, and* $\mathcal{A} \times W^c$ *is endowed with the product topology, then:*

(a) The following conditions are equivalent:

(a.i) The preference relation $\succcurlyeq^A$ *on* $\mathbb{C}$ *satisfies (A.1)–(A.8).*

(a.ii) There exists a continuous, monotonic increasing function $u^A : \mathbb{R} \to \mathbb{R}$*, a non-constant function* $v : \mathcal{A} \to \mathbb{R}$*, and a family of probability measures* $\{\pi^A(\cdot; a)\}_{a\in\mathcal{A}}$ *on* X *such that for all* $(a, w), (a', w') \in \mathbb{C}$*,* $(a, w) \succcurlyeq^A (a', w')$ *if and only if*

$$\sum_{x\in X} u^A(w(x))\,\pi^A(x; a) + v(a) \geq \sum_{x\in X} u^A(w'(x))\,\pi^A(x; a') + v(a').$$

(b) u *and* v *are unique up to positive linear transformation (that is, if* $\bar{u}^A$ *and* $\bar{v}$ *represent the preference relation* $\succcurlyeq^A$ *in the sense of (a.ii), then* $\bar{u}^A = \gamma\bar{u}^A + \zeta_u$*, and* $\bar{v} = \gamma v + \zeta_v$*, where* $\gamma > 0$*).*

(c) For each $a \in \mathcal{A}$*, given* $u^A(\cdot)$*,* $\pi^A(\cdot; a)$ *is unique and has full support.*

Remark: If the set of constant-valuation contracts, W^{cv}, is not the set of constant-payoff contracts (that is, $W^{cv} \neq W^c$) then the restricted hexagon condition can be stated for all $w, w', w'' \in W^{cv}$. The representation in Theorem 5 will become

$$(a, w) \mapsto \sum_{x\in X} \left[u^A(w(x))\,\sigma(x) + k(x)\right] \pi^A(x; a) + v(a),$$

where $\sigma(x) > 0$ and is unique and $\sum_{x\in X} k(x)\,\pi^A(x; a) = 0$ for all $a \in \mathcal{A}$.

In view of Theorem 5, if the agent's preference relation satisfies (A.1)–(A.6) then the ambiguity concerning the common beliefs assumption is no longer an issue. It amounts to the requirement that $\pi^P(\cdot \mid a) = \pi^A(\cdot \mid a)$, where $\pi^P(\cdot \mid a)$ and $\pi^A(\cdot \mid a)$ are given in Theorems 2 and 5, respectively.

5 Related Literature

Within the analytical framework of Anscombe and Aumann (1963), Drèze (1961), (1987) developed an axiomatic subjective expected utility theory with state-dependent preferences and moral hazard. Drèze argued, convincingly, that the "reversal of order" axiom of Anscombe and Aumann embodies the decision maker's belief that he may not influence the likely realization of alternative states. To allow for the possible presence of moral hazard Drèze relaxes the "reversal of order" axiom and assumes, instead, that the agent may strictly prefer knowing the outcome of a lottery (and, consequently, his payoff contingent on the state that obtains) before rather than after the state of nature is revealed. Presumably this is because knowing his contingent payoff ahead of time would allow the decision maker to take actions would tilt the probabilities of the states in his favor. In Drèze's theory the decision maker's actions are unobservable, and, consequently are tacit. The representation entails the maximization of subjective expected utility over a convex set of subjective probability measures. The set of probability measures represents the span of control the decision maker has over the likely realization of the states, presumably by his (implicit) choice of action.

The present work differs from that of Drèze in several important respects. First and most importantly, the analytical framework is different. In this paper the agent's actions constitute an explicit aspect of the choice set and the relation between the probability distributions on outcomes and the choice of action are explicit aspect of the representation. This difference represents distinct methodological outlooks. While Drèze is looking at the problem of modeling the agent's behavior from the perspective of an observer that may only observe the decision makers' preferences over acts and their preferences over the timing of resolution of the risk associated with the states generated by a lottery, I take the position that both the principal and the agent have well defined preferences over the set of action-contract pairs and can respond meaningfully to queries such as "How would you choose if confronted with two alternative action-contract pairs?" It is noteworthy that this type of questioning is endorsed by Savage:

> There is a mode of interrogation between what I called the behavioral and the direct. One can, namely, ask the person, not how he feels but what he would do in such and such situation. In so far as the theory of decision under development is regarded as an empirical one, the intermediate mode is a compromise between economy and rigor. But in the theory's more normative interpretation as a set of criteria of consistency for us to apply to our decisions, the intermediate mode is just the right one. (1972, p. 28).

Thus, even though the principal does not observe the agent's actions, these actions could be observed in principle, and in some situations also in practice.

Second, the approach developed in this paper is based on an analytical framework that makes it more readily and directly applicable to the parameterized distribution formulation of the principal-agent problem. Third, in the tradition of Savage (1954), the model of this paper does not invoke the notion of probabilities as a primitive concept.

6 Proofs

6.1 Preliminary results

The following two Lemmas establish results that will be used in the proofs below. The proofs of these lemmas appear in Karni (2004a).

Coordinate independence requires that, for every given action, the preference between any two contracts be independent of the payoffs assigned to outcomes on which the two contracts agree. This axiom is analogous to Savage's (1954) Sure Thing Principle; like the Sure Thing Principle coordinate independence implies the separability of the valuation of the monetary payoffs across outcomes.[6]

(**Coordinate independence**) *For all* $a \in \mathcal{A}$, $w, w' \in W$, $x \in X$, *and* $r, r' \in \mathbb{R}$,

$$(a, (w_{-x}, r)) \succcurlyeq \left(a, \left(w'_{-x}, r\right)\right) \text{ if and only if } (a, (w_{-x}, r')) \succcurlyeq \left(a, \left(w'_{-x}, r'\right)\right).$$

[6] Wakker (1989) provides a detailed discussion of coordinate independence and reviews the history of this axiom.

Lemma 6 *Let there be at least three nonnull outcomes. If* $\succcurlyeq$ *on* $\mathbb{C}$ *satisfies (A.3) then it satisfies coordinate independence.*

The Hexagon condition implies additive separable representation for actions that have exactly two nonnull outcomes**:**

(**Hexagon condition**) *For all* $a \in A$, $b \in B$, *and* $r, r', r'' \in \mathbb{R}$, *if* $X(a) = \{x, y\}$ *then,* $(a, (b_{-x}, r)_{-y}, r') \sim (a, (b_{-x}, r')_{-y}, r)$ *and* $(a, (b_{-x}, r)_{-y}, r'') \sim (a, (b_{-x}, r')_{-y}, r') \sim (a, (b_{-x}, r'')_{-y}, r)$ *imply* $(a, (b_{-x}, r')_{-y}, r'') \sim (a, (b_{-x}, r'')_{-y}, r')$.

Lemma 7 *Let there be exactly two nonnull outcomes. If* $\succcurlyeq$ *on* $\mathbb{C}$ *satisfies (A.3) then it satisfies the Hexagon condition.*

The following lemma asserts the existence jointly cardinal continuous additive value functions representing $\succcurlyeq^P$. It is an immediate corollary of Lemma 6 (and for the case in which $| X |= 2$, Lemma 7) and Wakker (1989) Theorem III.4.1.

Lemma 8 *Suppose that the full-support assumption holds. Then* $\succcurlyeq_a^h$ *is a continuous weak order on* W *satisfying (A.3) if and only if there exist jointly cardinal continuous additive value functions,* $\{V_a(\cdot; x)\}_{x \in X}$ *representing* $\succcurlyeq_a^h$ *on* W.

6.2 Proof of Lemma 1

$(i) \Rightarrow (ii)$. Suppose that $\succcurlyeq^h$ satisfies (A.1)-(A.3). By Lemma 8, there exist jointly cardinal continuous additive value functions, $\{V_a(\cdot;x)\}_{x\in X}$, representing $\succcurlyeq^h_a$ on W.

Claim: For all $a, a' \in \mathcal{A}$, $x \in X$, and $r', r \in \mathbb{R}$, $V_a(r';x) \geq V_a(r;x)$ if and only if $V_{a'}(r';x) \geq V_{a'}(r';x)$.

Proof of claim: Let $(a', (w_{-x}, r')) \succcurlyeq^h (a', (w_{-x}, r))$. But $(a', (w_{-x}, r')) \succcurlyeq^h (a', (w_{-x}, r'))$, $(a', (w_{-x}, r')) \succcurlyeq^h (a', (w_{-x}, r))$, and $(a, (w_{-x}, r')) \succcurlyeq^h (a, (w_{-x}, r'))$. Thus, by (A.3), $(a, (w_{-x}, r')) \succcurlyeq^h (a, (w_{-x}, r))$. The conclusion is implied by the representations of $\succcurlyeq^h$. ♣

Fix $\bar{a} \in \mathcal{A}$ then, for every $a, \in \mathcal{A}$ and $y \in X$, there exist $w, w', w'', w''' \in W$ such that

$$\sum_{x\in X-\{y\}} \left[V_{\bar{a}}(w(x), x) - V_{\bar{a}}(w'(x), x)\right] = \zeta > 0, \tag{9}$$

and

$$\sum_{x\in X-\{y\}} \left[V_a(w''(x), x) - V_a(w'''(x), x)\right] = \varepsilon > 0. \tag{10}$$

By continuity of $V_a(\cdot, x)$ and the connectedness of $\mathbb{R}$, for every $\hat{\zeta} \in [-\zeta, \zeta]$, $\hat{\varepsilon} \in [-\varepsilon, \varepsilon]$, and $y \in X$ there exist $w, w', w'', w''' \in W$ such that

$$\sum_{x\in X-\{y\}} \left[V_{\bar{a}}(w(x), x) - V_{\bar{a}}(w'(x), x)\right] = \hat{\zeta} \tag{11}$$

and

$$\sum_{x\in X-\{y\}} \left[V_a(w''(x), x) - V_a(w'''(x), x)\right] = \hat{\varepsilon}. \tag{12}$$

Define a family of functions $\{\phi_{(x;a)} : \mathbb{R} \to \mathbb{R} \,|a \in \mathcal{A}$ and $x \in X\}$ by $V_a(\cdot, x) = \phi_{(x;a)} V_{\bar{a}}(\cdot, x)$. Then $\phi_{(x;a)}$ is a continuous and, by Claim 1, nondecreasing function. To show that $\phi_{(x;a)}$ is positive affine function, fix $y \in X$ and let $I_y = V_{\bar{a}}(\mathbb{R}, y)$. Then, by the continuity of $V_{\bar{a}}(\cdot, y)$, I_y is an interval in $\mathbb{R}$. Take $\alpha, \beta, \gamma, \delta \in I_y$ such that $-\zeta \leq \alpha - \beta = \gamma - \delta \leq \zeta$ and $-\varepsilon \leq \phi_{(x;a)}(\alpha) - \phi_{(x;a)}(\beta) \leq \varepsilon$. Let $r, r', r'', r''' \in \mathbb{R}$ satisfy $V_{\bar{a}}(r, y) = \alpha$, $V_{\bar{a}}(r', y) = \beta$, $V_{\bar{a}}(r'', y) = \gamma$ and $V_{\bar{a}}(r''', y) = \delta$. Take $w, w' \in W$ such that

$$\sum_{x \in X - \{y\}} \left[V_{\bar{a}}(w(x), x) - V_{\bar{a}}(w'(x), x) \right] = \alpha - \beta. \tag{13}$$

Then, by the representation, $(\bar{a}, (w_{-y}; r)) \sim^h (\bar{a}, (w'_{-y}; r'))$ and $(\bar{a}, (w_{-y}; r'')) \sim^h (\bar{a}, (w'_{-y}; r'''))$.

Take $w'', w''' \in W$ such that

$$\sum_{x \in X - \{y\}} \left[V_a(w''(x), x) - V_a(w'''(x), x) \right] = \phi_{(x;a)}(\alpha) - \phi_{(x;a)}(\beta). \tag{14}$$

Since $V_a(\cdot, y) = \phi_{(y;a)} V_{\bar{a}}(\cdot, y)$ this implies $(a, (w''_{-y}, r)) \sim^h (a, (w'''_{-y}, r'))$. Applying (A.3) twice yields $(a, (w''_{-y}, r'')) \sim^h (a, (w'''_{-y}, r'''))$. Thus

$$\phi_{(y;a)}(\gamma) - \phi_{(y;a)}(\delta) = \sum_{x \in X - \{y\}} \left[V_a(w''(x), x) - V_a(w'''(x), x) \right] = \phi_{(y;a)}(\alpha) - \phi_{(y;a)}(\beta). \tag{15}$$

By Wakker (1987) Lemma 4.4 this implies that $\phi_{(y;a)}$ is affine. But y is nonnull hence, by Claim 1, $\phi_{(y;a)}$ is strictly increasing. Hence there exist $\lambda(y; a) > 0$ and $\xi(y; a)$ such that, for all $a \in \mathcal{A}$, $x \in X$, and $r \in \mathbb{R}$, $V_a(r, x) = \lambda(x; a) V_{\bar{a}}(r, x) + \xi(x; a)$. This completes the proof that $(i) \to (ii)$.

$(ii) \Rightarrow (i)$. That (ii) implies Axioms (A.1)–(A.2) follows from Lemma 8. To show that (ii) implies (A.3) assume that for all $a \in \mathcal{A}$ and $y \in X$ there exist positive affine transformations $\phi_{(y;a)}$ such that $V_a(\cdot, y) = \phi_{(y;a)} V_{\bar{a}}(\cdot, y)$. Suppose that $(\bar{a}, (w_{-y}, r)) \succcurlyeq^h (\bar{a}, (w'_{-y}, r'))$, $(\bar{a}, (w'_{-y}, r'')) \succcurlyeq^h (\bar{a}, (w_{-y}, r'''))$ and $(a, (w''_{-y}, r')) \succcurlyeq^h (a, (w'''_{-y}, r))$.

By the representation, $(\bar{a}, (w_{-y}, r)) \succcurlyeq^h (\bar{a}, (w'_{-y}, r'))$ if and only if

$$V_{\bar{a}}(r, y) + \sum_{x \in X - \{y\}} V_{\bar{a}}(w(x), x) \geq V_{\bar{a}}(r', y) + \sum_{x \in X - \{y\}} V_{\bar{a}}(w'(x), x), \tag{16}$$

and $(\bar{a}, (w'_{-y}, r'')) \succcurlyeq^h (\bar{a}, (w_{-y}, r'''))$ if and only if

$$V_{\bar{a}}(r''', y) + \sum_{x \in X - \{y\}} V_{\bar{a}}(w(x), x) \leq V_{\bar{a}}(r'', y) + \sum_{x \in X - \{y\}} V_{\bar{a}}(w(x), x). \tag{17}$$

Hence

$$V_{\bar{a}}(r', y) - V_{\bar{a}}(r, y) \leq \sum_{x \in X - \{y\}} [V_{\bar{a}}(w(x), x) - V_{\bar{a}}(w'(x), x)] \leq V_{\bar{a}}(r'', y) - V_{\bar{a}}(r''', y). \tag{18}$$

By positive affinity of $\phi_{(y;a)}$ these inequalities imply

$$V_{\bar{a}}(r', y) - V_{\bar{a}}(r, y) \leq V_{\bar{a}}(r'', y) - V_{\bar{a}}(r''', y). \tag{19}$$

Next observe that $(a, (w''_{-y}, r')) \succcurlyeq^P (a, (w'''_{-y}, r))$ if and only if

$$V_a(r', y) + \sum_{x \in X - \{y\}} V_a(w''(x), x) \geq V_a(r, y) + \sum_{x \in X - \{y\}} V_a(w'''(x), x). \tag{20}$$

Thus

$$V_a(r', y) - V_a(r, y) \geq \sum_{x \in X - \{y\}} [V_a(w'''(x), x) - V_a(w''(x), x)]. \tag{21}$$

Hence inequality (19) implies

$$V_a(r'', y) + \sum_{x \in X - \{y\}} V_a(w''(x), x) \geq V_a(r''', y) + \sum_{x \in X - \{y\}} V_a(w'''(x), x). \tag{22}$$

Hence $(a, (w''_{-y}, r'')) \succcurlyeq^h (a, (w'''_{-y}, r'''))$. Thus $(ii) \to (i)$. ■

6.3 Proof of Theorem 2

$(a.i) \rightarrow (a.ii)$. By Lemma 8, for each $a \in \mathcal{A}$, $\succcurlyeq_a^P$ on W there exist jointly cardinal continuous additive value functions, $\{V_a(\cdot; x)\}_{x \in X}$, representing $\succcurlyeq_a^P$. Fix $\bar{a} \in \mathcal{A}$ then, by Lemma 1, there exist numbers $\lambda(x,a) > 0$ and $\xi(x,a)$ such that $V_a(\cdot; x) = \lambda(x,a) V_{\bar{a}}(\cdot; x) + \xi(x,a)$ for all $x \in X$ and $a \in A$.

Let w^{**} and w^* be constant valuation contracts satisfying $w^{**} \succ^P w^*$. Invoking the uniqueness of the jointly cardinal additive value functions, set $V_a(w^*(x), x) = 0$, for all $x \in X$, $a \in \mathcal{A}$, and $\sum_{x \in X} V_a(w^{**}(x), x) = 1$ for all $a \in \mathcal{A}$. Then

$$0 = V_a(w^*(x), x) = \lambda(x,a) V_{\bar{a}}(w^*(x), x) + \xi(x,a) = \xi(x,a), \text{ for all } a \in \mathcal{A}. \tag{23}$$

Fix $\eta \in \mathbb{R}$ and let w^η satisfy $V_{\bar{a}}(w^\eta(x), x) = \eta V_{\bar{a}}(w^{**}(x), x)$ for all $x \in X$. Then $V_a(w^\eta(x), x) = \lambda(x,a) V_{\bar{a}}(w^\eta(x), x) = \lambda(x,a) \eta V_{\bar{a}}(w^{**}(x), x) = \eta V_a(w^{**}(x), x)$. Hence, by definition of constant-valuation contracts, for all $a \in A$,

$$\sum_{x \in X} V_a(w^\eta(x), x) = \eta = \sum_{x \in X} V_{\bar{a}}(w^\eta(x), x). \tag{24}$$

Let $(a, w) \succcurlyeq^P (a', w')$ then, by (A.2), there is a constant valuation contract, $\tilde{w}$, such that $(a, w) \succcurlyeq^P (a, \tilde{w}) \sim^P (a', \tilde{w}) \succcurlyeq^P (a', w')$. Hence

$$(a, \tilde{w}) \sim^P (a', \tilde{w}) \Leftrightarrow \sum_{x \in X} V_a(\tilde{w}(x), x) = \sum_{x \in X} V_{a'}(\tilde{w}(x), x). \tag{25}$$

By transitivity,

$$(a, w) \succcurlyeq^P (a', w') \Leftrightarrow \sum_{x \in X} V_a(w(x), x) \geq \sum_{x \in X} V_{a'}(w'(x), x). \tag{26}$$

Thus $\{V_a\}_{a\in\mathcal{A}}$ represent $\succcurlyeq^{P.}$

For every $r \in \mathbb{R}$ and $x \in X$ let $\left(w^r_{-x}, r\right) \in W$ be a constant-valuation contract. Define $u^P(r, x; a) := \sum_{y\in X-\{x\}} V_a(w^r(y), y) + V_a(r, x)$, $a \in A$. But for any constant-valuation contract, $\tilde{w}$, $\sum_{x\in X} V_a(\tilde{w}(x), x) = \sum_{x\in X} V_{a'}(\tilde{w}(x), x)$, for all $a, a' \in A$. Thus $u^P(r, x; a) = u^P(r; x)$ for all $a \in A$.

Define $\pi^P(x; a) := V_a(w^{**}(x), x)$, for all $x \in X$ and $a \in \mathcal{A}$. But $V_a(r, x) = \rho V_a(w^{**}(x), x)$. Hence

$$u^P(r; x) = \sum_{y\in X-\{x\}} V_a(w^r(y), y) + V_a(r, x) = \rho \sum_{y\in X} V_a(w^{**}(y), y).$$

But

$$\frac{V_a(r, x)}{u^P(r, x)} = \frac{\rho V_a(w^{**}(x), x)}{\rho \sum_{y\in X} V_a(w^{**}(y), y)} = V_a(w^{**}(x), x) = \pi^P(x; a).$$

Thus $u^P(r; x)\pi^P(x; a) = V_a(r; x)$, for all $r \in \mathbb{R}$, $x \in X$ and $a \in \mathcal{A}$. Equation (26) implies that $\succcurlyeq^P$ is represented by $(a, w) \longmapsto \sum_{x\in X} u^P(w(x), x)\pi^P(x; a)$. This completes the proof that $(a.i) \to (a.ii)$.

$(a.ii) \to (a.i)$. The proof that $(a.ii)$ implies (A.1) and (A.2) is immediate. The proof that it implies (A.3) follows from Lemma 1.

(b) The uniqueness of the jointly cardinal additive representation implies that, for all $a \in A$ and $x \in X$, $v^P(\cdot; x)\pi^P(x; a) = \lambda(a) u^P(\cdot, x)\pi^P(x; a) + \zeta(x; a)$, $\lambda(a) > 0$. Thus $\sum_{x\in X(a)} v^P(w^*(x); x)\pi^P(x; a) = \sum_{x\in X(a)} \zeta(x; a) = \sum_{x\in X(a)} \frac{\zeta(x;a)}{\pi^P(x;a)}\pi^P(x; a)$, for all $a \in A$. Consequently, $\sum_{x\in X(a)} \zeta(x; a) = c$ is constant. Thus, $\zeta(x; a)/\pi^P(x; a) = c$. Hence, for all

$a, a' \in A,$

$$\sum_{x \in X(a)} v^P \left(w^{**}(x); x\right) \left[\pi(x; a) - \pi(x; a')\right] = \lambda(a) - \lambda(a').$$

But w^{**} is a constant valuation contract, hence $\sum_{x \in X(a)} v^P \left(w^{**}(x); x\right) \left[\pi(x; a) - \pi(x; a')\right] = 0$ for all $a, a' \in A$. Thus $\lambda(a) = \lambda(a') = \lambda > 0$ for all $a, a' \in A$, and $v^P(\cdot; x) = \lambda u^P(\cdot, x) + c$.

(c) To prove the uniqueness of $\{\pi^P(\cdot; a)\}_{a \in \mathcal{A}}$ suppose that for some $a \in \mathcal{A}$ there exist another probability measure $\mu(\cdot; a)$ such that, for some $x, x' \in X$, $\pi^P(x; a) > \mu(x; a)$ and $\pi^P(x'; a) < \mu(x'; a)$ and $(u_\mu; \mu)$ represent the $\succcurlyeq_a^P$. This implies that $u^P(r; z) \pi^P(z; a) = u_\mu(r; z) \mu(z; a), z \in \{x, x'\}$. Hence $u^P(r; x) < u_\mu(r, x)$ and $u^P(r; x') > u_\mu(r, x')$ for all $r \in \mathbb{R}$. Let $\tilde{w}$ be a constant valuation contract. Then, by definition, $u^P(\tilde{w}(x), x) = u^P(\tilde{w}(x'), x')$. Thus $u_\mu(\tilde{w}(x), x) > u_\mu(\tilde{w}(x'), x')$ a contradiction. The fact that for each $a \in \mathcal{A}$, $\pi^P(\cdot; a)$ has full support is an implication of the full-support assumption. ■

6.4 Proof of Theorem 4

$(a.i) \Rightarrow (a.ii)$. By Lemma 3, for each $a \in \mathcal{A}$, there exist continuous real-valued function, $U^A(\cdot; a)$, on $\mathbb{R}$, and a full support probability measure, $p^A(\cdot; a)$, on X such that $\sum_{x \in X} U^A(w(x), a) p^A(x; a)$ represents of $\succcurlyeq_a^A$ on W, where U^A is unique up to positive affine transformations.

Fix $\bar{a} \in A$ then, by (A.3) and Lemma 1, the relation $\succcurlyeq^A$ on $\mathbb{C}$ satisfies (A.1)–(A.4) if and only if there exist $\lambda(x;a) > 0$ and $\xi(x;a)$ such that, for all $(x;a) \in X \times \mathcal{A}$,

$$U^A(\cdot, a)\, p^A(x;a) = \lambda(x;a)\, U^A(\cdot;\bar{a})\, p^A(x;\bar{a}) + \xi(x;a).$$

Hence, for every $a \in \mathcal{A}$, $\sum_{x \in X} \lambda(x;a)\, p^A(x;\bar{a}) = b(a) > 0$ and $\sum_{x \in X} \xi(x;a) = v(a)$, for all $x \in X$. Moreover, by (A.5), for every $a \in \mathcal{A}$, $U^A(\cdot, a)$ is monotonic increasing.

Define $u^A(r) := U^A(r, \bar{a})$ then, for every $a \in \mathcal{A}$ and $r \in \mathbb{R}$, $U^A(r,a) = b(a)\, u^A(r) + v(a)$. Let r^{**}, $r^* \in \mathbb{R}$ satisfy $u^A(r^{**}) > u^A(r^*)$. Invoking the uniqueness of $U^A(\cdot, \bar{a})$ (see Lemma 3) set $u^A(r^{**}) = 1$ and $u^A(r^*) = 0$.

Let $W^c \subset W$ be the subset of constant-payoff contracts (that is, W^c is the subset of constant functions in W). With slight abuse of notation let r denote the contract in W^c whose payoff is r regardless of the outcome. For every $a \in \mathcal{A}$ and $r' \in W^c$ define $r(a, r') \in W^c$ by $(a, r(a, r')) \sim^A (\bar{a}, r')$. By (A.2) and (A.5) $r(a, r')$ exists and is unique, thus $r(a, r')$ is well-defined. Moreover,

$$U^A(r(a, r^{**}), a) = b(a) + v(a) \tag{27}$$

and

$$U^A(r(a, r^*), a) = v(a). \tag{28}$$

Thus, for all $a, a' \in A$ and $r' \in W^c$, $(a, r(a, r')) \sim^A (a', r(a', r'))$ if and only if

$$b(a)\, u^A(r(a, r')) + v(a) = u^A(r') = b(a')\, u^A(r(a', r')) + v(a'). \tag{29}$$

Let $(a, w) \succcurlyeq^A (a', w')$ then, by continuity, there is $\hat{r} \in W^c$ such that $(a, w) \succcurlyeq^A (a, r(a, \hat{r})) \sim^A$

$(a', r(a', \hat{r})) \succcurlyeq^A (a', w')$. But, by equation (29), $(a, r(a, \hat{r})) \sim^A (a', r(a', \hat{r}))$ if and only if

$$b(a) u^A (r(a, \hat{r})) + v(a) = b(a') u^A (r(a', \hat{r})) + v(a'). \tag{30}$$

Hence, by transitivity, $(a, w) \succcurlyeq^A (a', w')$ if and only if

$$\sum_{x \in X} b(a) u^A (w(x)) p^A (x; a) + v(a) \geq \sum_{x \in X} b(a') u^A (w'(x)) p^A (x; a') + v(a'). \tag{31}$$

Hence $(a.i) \rightarrow (a.ii)$.

The proof that $(a.ii) \rightarrow (a.i)$ is immediate. The proof of parts (b) and (c) follow from the corresponding parts of Lemma 8 . ■

6.5 Proof of Theorem 5

$(i) \rightarrow (ii)$. The preference relation $\succcurlyeq^A$ restricted to $\mathcal{A} \times W^c$ is a continuous weak order satisfying coordinate independence, (A.7), and, by (A.8), both $\mathcal{A}$ and W^c are essential with respect to $\succcurlyeq^A$. It also satisfies the restriction hexagon condition (A.6). Hence, by Theorem III.4.1 of Wakker (1989), there exist jointly cardinal continuous additive valued functions $V_1 : \mathcal{A} \rightarrow \mathbb{R}$ and $V_2 : W^c \rightarrow \mathbb{R}$ representing $\succcurlyeq^A$ (that is, for all $(a, w), (a', w') \in \mathcal{A} \times W^c$, $(a, w) \succcurlyeq^A (a', w')$ if and only if $V_1(a) + V_2(w) \geq V_1(a') + V_2(w')$. By Theorem 4, Axioms (A.1)–(A.5) imply that $(a, w) \succcurlyeq^A (a', w')$ if and only if

$$b(a) u^A (w) + v(a) \geq b(a') u^A (w') + v(a'). \tag{32}$$

By (A.8), v is a non-constant function.

Take $a, a', a'' \in \mathcal{A}$ and $w, w', w'' \in W^c$ that satisfy (A.6). Then, by Theorem 4, (A.6) implies that

$$b(a)\,u^A(w') + v(a) = b(a')\,u^A(w) + v(a'), \tag{33}$$

$$b(a)\,u^A(w'') + v(a) = b(a')\,u^A(w') + v(a') = b(a'')\,u^A(w) + v(a''), \tag{34}$$

and

$$b(a')\,u^A(w'') + v(a') = b(a'')\,u^A(w') + v(a''). \tag{35}$$

Equations (33), (34) and (35) imply that

$$\kappa = v(a') - v(a) + u^A(w')\,[b(a') - b(a)] = v(a'') - v(a') + u^A(w)\,[b(a'') - b(a')] \tag{36}$$

and

$$\kappa' = v(a') - v(a) + u^A(w'')\,[b(a') - b(a)] = v(a'') - v(a') + u^A(w')\,[b(a'') - b(a')], \tag{37}$$

where $\kappa = b(a')\,[u^A(w') - u^A(w)]$ and $\kappa' = b(a')\,[u^A(w'') - u^A(w')]$. But $b(a')\,[u^A(w') - u^A(w)] = \gamma\,[V_2(w') - V_2(w)]$ and $b(a')\,[u^A(w'') - u^A(w')] = \gamma\,[V_2(w'') - V_2(w')]$, where $\gamma > 0$. Moreover,

$$V_2(w') - V_2(w) = V_1(a') - V_1(a) = V_2(w'') - V_2(w').$$

Thus $\kappa = \kappa'$ (that is, $u^A(w') - u^A(w) = u^A(w'') - u^A(w')$). Equations (36) and (37) imply that

$$[b(a') - b(a)]\left[u^A(w'') - u^A(w')\right] = 0 = [b(a'') - b(a')]\left[u^A(w') - u^A(w)\right]. \tag{38}$$

But $u^A(w'') - u^A(w') = u^A(w') - u^A(w) \neq 0$. Hence

$$b(a) = b(a') = b(a''). \tag{39}$$

Thus $\sum_{x \in X} u^A(w) p^A(x; a)$ is independent of a. Define $\pi^A(x; a) = p^A(x; a)$ for all $x \in X$ and $a \in \mathcal{A}$. This implies the representation in Theorem 5.[7]

That the representation $(a, w) \mapsto \sum_{x \in X} u^A(w(x)) \pi^A(x; a) + v(a)$ implies (A.6)-(A.8) is immediate. That it implies (A.1)-(A.5) follows from Theorem 4. ■

References

[1] Anscombe, Francis J., and Robert J. Aumann (1963) "A Definition of Subjective-Probability," *Annals of Mathematical Statistics* 34: 199-205.

[2] Chambers, Robert, J., and John Quiggin (2000) *Uncertainty, Production, Choice, and Agency.* Cambridge, Cambridge University Press.

[3] Drèze, Jacques H. (1961) "Les fondements logiques de l'utilite cardinale et de la probabilite subjective," *La Decision,* Colloques Internationaux de CNRS, 73-87.

[4] Drèze, Jacques H. (1987) "Decision Theory with Moral Hazard and State-Dependent Preferences," In Drèze, Jacques. H. *Essays on Economic Decisions Under Uncertainty.* Cambridge: Cambridge University Press.

[7] The change of notation form p^A to π^A was introduced to emphasize that these are probabilities that represent the agent's beliefs when the utility functions are outcome independent.

[5] Grossman, Sanford J., and Oliver D. Hart (1983) "An Analysis of the Principal-Agent Problem," *Econometrica* 51: 7-45.

[6] Hart, Oliver, and Bengt Holmstrom (1987) "The Theory of Contracts," in Bewley, Truman (ed.) *Advances in Economic Theory.* Cambridge: Cambridge University Press.

[7] Karni, Edi (1993) "A Definition of Subjective Probabilities with State-Dependent Preferences," *Econometrica* 61: 187-198.

[8] Karni, Edi (1996) "Probabilities and Beliefs," *Journal of Risk and Uncertainty* 13: 249-262.

[9] Karni, Edi (2003) "On the Representation of Beliefs by Probabilities," *Journal of Risk and Uncertainty* 26: 17-38.

[10] Karni, Edi (2004) "Subjective Expected utility Theory with Costly Actions," *Games and Economic Behavior* (forthcoming).

[11] Karni, Edi (2004a) "Subjective Expected Utility Theory without States of the World," Working paper, JHU.

[12] Karni, Edi, and David Schmeidler (1993) "On the Uniqueness of Subjective Probabilities," *Economic Theory* 3: 267-277.

[13] Mirrlees, James (1974) "Notes on Welfare Economics, Information and Uncertainty," In M. Balch, D. McFadden, and S. Wu (eds.) *Essays in Economic Behavior Under Uncertainty.* Amsterdam: North Holland.

[14] Mirrlees, James (1976) "The Optimal Structure of Authority and Incentives Within an Organization," *The Bell Journal of Economics 7: 105-131.*

[15] Salanié, Bernard (1997) *The Economics of Contracts.* Cambridge: The MIT Press.

[16] Savage, Leonard, J. (1954) *The Foundations of Statistics.* New York: John Wiley and Sons. Second revised edition 1972.

[17] Schervish, Mark, J., Teddy Seidenfeldt, and Joseph B. Kadane (1990) "State-Dependent Utilities," *Journal of American Statistical Association,* 85: 840-847.

[18] Wakker, Peter, P. (1987) "Subjective Probabilities for State-Dependent Continuous Utility," *Mathematical Social Sciences* 14: 289-298.

[19] Wakker, Peter, P. (1989) *The Additive Representations of Preferences.* Dordrecht: Kluwer Academic Publishers.