האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

BAYESIAN DECISION THEORY AND THE REPRESENTATION OF BELIEFS

by

EDI KARNI

Discussion Paper # 444 January 2007

מרכז לחקר הרציונליות

CENTER FOR THE STUDY
OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Bayesian Decision Theory and the Representation of Beliefs

Edi Karni*

Johns Hopkins University

January 15, 2007

Abstract

In this paper, I present a Bayesian decision theory and define choice-based subjective probabilities that faithfully represent Bayesian decision makers' prior and posterior beliefs regarding the likelihood of the possible effects contingent on his actions. I argue that no equivalent results can be obtained in Savage's (1954) subjective expected utility theory and give an example illustrating the potential harm caused by ascribing to a decision maker subjective probabilities that do not represent his beliefs.

*I am grateful to Itzhak Gilboa, Simon Grant, Robert Nau, Tiemen Woutersen, Eddie Dekel and, especially, Peter Wakker for their comments and to the NSF for financial support under grant SES-0314249. Part of the work reported here was done during my visit to the Center for the study of rationality and interactive decisions, at the Hebrew University of Jerusalem.

1 Introduction

Decision making under uncertainty involves choosing among courses of action whose exact consequences are not determined solely by the decision maker's choice. With few exceptions, modeling decision making under uncertainty invoke the analytical framework of Savage (1954), consisting of a set of states, S; a set of consequences, C; and the set, F, of all the mappings from the set of states to the set of consequences. Elements of F, referred to as *acts*, correspond to conceivable courses of action; a decision maker is characterized by a preference relation on F. States resolve the uncertainty in the sense that once the (unique) true state becomes known, the unique consequence implied by each and every act becomes known. Subsets of S are *events*. An event is said to obtain if the true state belongs to it.

In Savage's subjective expected utility theory, the structure of a (prior) preference relation, $\succcurlyeq$, on F, depicted axiomatically, allows its representation by an expected utility functional,

$$\int\limits_S u\left(f(s)\right) d\pi\left(s\right), \tag{1}$$

where u is a real-valued (utility) function defined on the consequences; π is a finitely additive, nonatomic probability measure on S; and $f \in F$. Moreover, the utility function u is unique up to positive linear transformation, and, given u, the subjective probability measure π is unique. Presumably, the interest in this representation stems from the notion that the decision maker evaluates alternative acts by assessing, separately, the merit and likelihood of the possible consequences and then integrates these assessments. In other words, the merit of the consequences and the decision maker's degree of belief regarding their likelihood are meaningful cognitive phenomena that may be quantified, respectively, by a utility function and a subjective probability measure. These numerical values are combined, using the functional in equation (1), to obtain a numerical valuation of the act. In this interpretation, a decision maker's beliefs is a relation on the set of events whose meaning is "at least as likely to obtain as."

Clearly, any probability measure on S defines beliefs.[1] In Savage's (1954) theory the decision maker's (prior) beliefs are defined by the probability measure π. However, the uniqueness of π and, consequently, of Savage's beliefs, is predicated on the convention that constant acts are constant utility acts (that is, the utility function is state-independent). *This convention, however, is not implied by the axioms and, more important, is not testable within Savage's analytical framework.*[2] Yet without it, it is impossible to disentangle the probabilities and marginal utilities. Consequently, there are infinitely many prior probability measures consistent with a decision maker's prior preferences. In other words, even if a decision maker's beliefs constitute a quantifiable psychological phenomenon and his choice behavior is consistent with the axiomatic structure of expected utility theory, the proposition that the subjective probabilities ascribed to him by Savage's model represent the decision maker's beliefs is untestable. In what follows I use the term *identifiable* subjective probabilities to refer to a probability measure, consistent with the decision maker's choice behavior, whose

[1] Let $\succeq_{\mathcal{B}}$ be a binary relation on the set of events defined by $E \succeq_{\mathcal{B}} E'$ if and only if $\pi(E) \geq \pi(E')$, for all events E and E'.

[2] For more details regarding the first part of this assertion, see Karni (1996, 2006).

uniqueness is not predicated on a particular choice of a utility function. The argument above implies that Savages' subjective probability measure is unidentifiable.

A decision maker who, upon receiving new information, displays a change of preferences reflecting his modified beliefs is said to be Bayesian if the representation of his prior preferences involve an identifiable prior probability measure on the set S; the representations of his posterior preference relations involve well-defined posterior probability measures on S; and the representations of the posterior preference relations are obtained from the prior preference relation by updating the prior probabilities according to Bayes rule. Next I show that Savage's model fails to yield an identifiable prior and is not a model of Bayesian decision making.[3]

To prove this assertion, let γ be a strictly positive, bounded, real-valued function on S, and let $E(\gamma) = \int_S \gamma(s)\, d\pi(s)$. Then the prior preference relation, depicted by the representation (1) is also represented by

$$\int_S \hat{u}(f(s), s)\, d\hat{\pi}(s), \tag{2}$$

where $\hat{u}(\cdot, s) = u(\cdot)/\gamma(s)$ and $\hat{\pi}(s) = \pi(s)\gamma(s)/E(\gamma)$, for all $s \in S$.[4]

Consider next an experiment whose possible outcomes are pertinent to the decision maker's assessment of the likely realization of events. Let X be the set of observations, and

[3] The same criticism applies to all subjective expected utility models that invoke Savage's analytical framework, such as the model of Anscombe and Aumann (1963).

[4] This point has been recognized by Drèze (1987); Schervish, Seidenfeldt, and Kadane (1990); Karni (1996), (2003); Karni and Schmeidler (1993); and Nau (1995).

denote by $q\left(x \mid s\right)$ the conditional probability of the observation x if the true state is s (s may be thought of as the parameters underlying the distribution of X) and $\{q\left(\cdot \mid s\right) \mid s \in S\}$ is a family of likelihood functions. Then, invoking representation (1), the induced posterior preference relations $\{\succcurlyeq^x \mid x \in X\}$ of a Bayesian decision maker are defined as follows: for all $x \in X$, and $f, f' \in F$,

$$f \succcurlyeq^x f' \Leftrightarrow \int_S u\left(f(s)\right) d\pi\left(s \mid x\right) \geq \int_S u\left(f'(s)\right) d\pi\left(s \mid x\right), \tag{3}$$

where, for each $x \in X$,

$$\pi\left(\cdot \mid x\right) = \frac{q\left(x \mid \cdot\right)\pi\left(\cdot\right)}{\displaystyle\int_S q\left(x \mid s'\right) d\pi\left(s'\right)} \tag{4}$$

is the posterior probability measure obtained by the application of Bayes' theorem.

Invoking representation (2), define the induced posterior preference relations, $\{\hat{\succcurlyeq}^x \mid x \in X\}$, of the same Bayesian decision maker whose prior probability measure is $\hat{\pi}\left(\cdot\right)$ as follows: for all $x \in X$, and $f, f' \in F$,

$$f \hat{\succcurlyeq}^x f' \Leftrightarrow \int_S \hat{u}\left(f(s), s\right) d\hat{\pi}\left(s \mid x\right) \geq \int_S \hat{u}\left(f'(s), s\right) d\hat{\pi}\left(s \mid x\right), \tag{5}$$

where

$$\hat{\pi}\left(\cdot \mid x\right) = \frac{q\left(x \mid \cdot\right)\hat{\pi}\left(\cdot\right)}{\displaystyle\int_S q\left(x \mid s'\right) d\hat{\pi}\left(s'\right)}. \tag{6}$$

However,

$$\int_S \hat{u}\left(f(s), s\right) d\hat{\pi}\left(s \mid x\right) = \left[\frac{\displaystyle\int_S q\left(x \mid s'\right) d\pi\left(s'\right)}{\displaystyle\int_S q\left(x \mid s'\right) d\hat{\pi}\left(s'\right)}\right] \int_S u\left(f(s)\right) d\pi\left(s \mid x\right). \tag{7}$$

Hence for all $f, f' \in F$ and $x \in X$,

$$\int_S \hat{u}\left(f(s), s\right) d\hat{\pi}\left(s \mid x\right) \geq \int_S \hat{u}\left(f'(s), s\right) d\hat{\pi}\left(s \mid x\right) \tag{8}$$

if and only if

$$\int_S u\left(f(s)\right) d\pi\left(s \mid x\right) \geq \int_S u\left(f'(s)\right) d\pi\left(s \mid x\right). \tag{9}$$

Thus $\succcurlyeq^x = \hat{\succcurlyeq}^x$ for all $x \in X$. *The fact that a decision maker updates his preferences using Bayes' rule does not imply that either his prior or his posterior beliefs, as defined by the representing probabilities, are unique.*

I presented an alternative framework for the analysis of decision making under uncertainty in Karni (2006). In this paper I use the same analytical framework to develop a theory of decision making under uncertainty in which the prior and posterior subjective probabilities of Bayesian decision makers are identifiable. The identifiability of the decision maker's beliefs and the definition of subjective probabilities is not a purely philosophical issue - it has potential economic consequences. In the appendix I analyze a moral hazard problem illustrating the potential pitfalls of ascribing the agent an incorrect prior. In that example the principal knows the agent's prior preference relation and uses Bayes' rule to update the agent's preferences in the light of new information. However, even though the agent is Bayesian, because the principal ascribes to him an incorrect prior, the principal fails to design an incentive compatible contract.

Section 2 reviews the analytical framework and states the main result of Karni (2006).

Section 3 presents a Bayesian decision theory and shows that the prior and posterior subjective probabilities are identifiable. Concluding remarks appear in section 4.

2 The Theory

2.1 The analytical framework

Let Θ be a finite set of *effects* and A a set of *actions.* Actions represent initiatives that decision makers may take to influence the likely realization of effects. A *bet*, b, is a mapping from Θ into $\mathbb{R}$, the set of real numbers. Bets have the interpretation of monetary payoffs contingent on effects. Let $B := \mathbb{R}^{\Theta}$ denote the set of all bets, and assume that it is endowed with the $\mathbb{R}^{|\Theta|}$ topology. Denote by $(b_{-\theta}r)$ the bet obtained from $b \in B$ by replacing the θ coordinate of b (that is, $b(\theta)$) with r. In this framework, effects are analogous to states, in the sense that they resolve the uncertainty associated with bets. However, unlike states, decision makers believe that they may influence the likely realization of effects by their actions.

Decision makers are supposed to choose actions and, at the same time, place bets on the effects. For example, a decision maker may adopt an exercise and diet regimen to reduce the risk of heart attack and simultaneously take out health insurance and life insurance policies. The diet and exercise regimens correspond to actions, the states of health are effects, and the financial terms of an insurance policy constitute a bet. Formally, the *choice set,* $\mathbb{C}$, consists of all action-bet pairs (that is, $\mathbb{C} = A \times B$). A choice of an action a and a bet b results in an effect-payoff pair, or *consequence*, $(\theta, b(\theta))$. Let C denote the set of all consequences $(C = \Theta \times \mathbb{R})$.

A decision maker is characterized by prior preference relation, $\succcurlyeq$, on $\mathbb{C}$. The strict preference relation, $\succ$, and the indifference relation, $\sim$, are the asymmetric and symmetric parts of $\succcurlyeq$, respectively. For each $a \in A$, the preference relation $\succcurlyeq$ on $\mathbb{C}$ induces a conditional preference relation on B defined as follows: for all $b, b' \in B$, $b \succcurlyeq_a b'$ if and only if $(a, b) \succcurlyeq (a, b')$.

An effect, θ, is said to be *nonnull given the action* a if $(a, (b_{-\theta} r)) \succ (a, (b_{-\theta} r'))$, for some $b \in B$ and $r, r' \in \mathbb{R}$; it is *null given the action* a otherwise. I assume that every effect is nonnull for some action a. Given a preference relation, $\succcurlyeq$, denote by $\Theta(a; \succcurlyeq)$ the subset of effects that are nonnull given a according to $\succcurlyeq$. To simplify the notation, when there is no risk of confusion, I write $\Theta(a)$ instead of $\Theta(a; \succcurlyeq)$.

Two effects, θ and θ', are said to be *elementarily linked* if there are actions $a, a' \in A$ such that $\theta, \theta' \in \Theta(a) \cap \Theta(a')$. Two effects are said to be *linked* if there exists a sequence of effects $\theta = \theta_0, ..., \theta_n = \theta'$ such that every θ_j is elementarily linked with θ_{j+1}. The condition that every pair of effects be linked is used, below, to join the utility scales associated with the different actions.

Actions may affect the decision maker's well-being directly. For example, adopting a diet regimen may require that the decision maker avoid eating food he likes. With this in mind, I refer to a bet as constant valuation if, once accepted, it leaves the decision maker indifferent among actions whose direct utility costs are just compensated for by the improved chances of winning better outcomes they afford.[5] I assume that the set A consists of those actions whose induced variations in the likely realization of alternative effects and associated direct utility implications determine a unique set of constant valuation bets. Formally, let $I(a;b) = \{b' \in B \mid (a,b') \sim (a,b)\}$. The idea of constant valuation bets is then formalized as follows:

Definition 1: A bet $\bar{b}$ is said to be a *constant-valuation bet* if $\left(a,\bar{b}\right) \sim \left(a',\bar{b}\right)$ for all $a, a' \in A$, and $\cap_{a \in A} I\left(a;\bar{b}\right) = \{\bar{b}\}$.

The last requirement implies that if $\bar{b}$ is a constant valuation bet, no other $b \in B$ satisfies $(a,b) \sim (a',b)$ for all $a, a' \in A$ and $(a,b) \sim \left(a,\bar{b}\right)$ for some $a \in A$.[6] Let B^{cv} denote the set of all constant valuation bets. If b^{**} and b^{*} are constant valuation bets satisfying $(a',b^{**}) \succ (a',b^{*})$ for some $a' \in A$, then, by transitivity of $\succcurlyeq$, $(a,b^{**}) \succ (a,b^{*})$ for all $a \in A$. Since transitivity will be assumed, I write $b^{**} \succ b^{*}$ instead of $(a,b^{**}) \succ (a,b^{*})$.

The following richness assumption is maintained throughout:

(A.0) *Every pair of effects is linked, there exist constant-valuation bets b^{**} and b^{*} such that $b^{**} \succ b^{*}$ and, for every $(a,b) \in \mathbb{C}$, there is $\bar{b} \in B^{cv}$ satisfying $(a,b) \sim \bar{b}$.*

[5] Because of the possible direct impact of the actions on the decision maker's well-being, constant valuation bets are different from Drèze's (1987) notion of "omnipotent" acts and the idea of constant valuation acts in Karni (1993, 2006) and Skiadas (1997).

[6] The model may be extended to include additional actions (see in Karni 2006).

2.2 Axioms and Representation

The first two axioms are standard.

(A.1) (**Weak order**) $\succcurlyeq$ *on* $\mathbb{C}$ *is a complete and transitive binary relation.*

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

The third axiom requires that the "intensity of preferences" for monetary payoffs contingent on any given effect be independent of the action that resulted in that effect.

(A.3) (**Action-independent betting preferences**) *For all* $a,a'\in A$, $b,b',b'',b'''\in B$, $\theta\in\Theta(a)\cap\Theta(a')$, *and* $r,r',r'',r'''\in\mathbb{R}$, *if* $(a,b_{-\theta}r)\succcurlyeq\left(a,b'_{-\theta}r'\right)$, $\left(a,b'_{-\theta}r''\right)\succcurlyeq(a,b_{-\theta}r''')$, *and* $\left(a',b''_{-\theta}r'\right)\succcurlyeq\left(a',b'''_{-\theta}r\right)$ *then* $\left(a',b''_{-\theta}r''\right)\succcurlyeq\left(a',b'''_{-\theta}r'''\right)$.

To grasp the meaning of action-independent betting preferences, think of the preferences $(a,b_{-\theta}r)\succcurlyeq\left(a,b'_{-\theta}r'\right)$ and $\left(a,b'_{-\theta}r''\right)\succcurlyeq(a,b_{-\theta}r''')$ as indicating that, given an action a and an effect θ, the intensity of the preferences of r'' over r''' is sufficiently larger than that of r over r' as to reverse the preference ordering of the effect-contingent payoffs $b_{-\theta}$ and $b'_{-\theta}$. The axiom requires that these intensities not be contradicted when the action is a' instead of a and bets are b'' and b'''.[7]

[7] Action-independent betting preferences invokes Wakker's (1987) idea of cardinal consistency. A more elaborate discussion of this axiom is provided in Karni (2006).

Karni (2006) shows that a preference relation on $\mathbb{C}$ satisfying (A.0), has the structure described by axioms (A.1)–(A.3) if and only if there exist a continuous utility function, u, on the set of consequences, a family of action-dependent probability measures, $\{\pi(\cdot; a) \mid a \in A\}$, on the set of effects, and a family of continuous strictly increasing functions $\{f_a : \mathbb{R} \to \mathbb{R} \mid a \in A\}$ such that, for all $(a, b), (a', b') \in \mathbb{C}$,

$$(a, b) \succcurlyeq (a', b') \Leftrightarrow f_a\left(\sum_{\theta\in\Theta} u(b(\theta); \theta)\, \pi(\theta; a)\right) \geq f_{a'}\left(\sum_{\theta\in\Theta} u(b'(\theta); \theta)\, \pi(\theta; a')\right). \tag{10}$$

Furthermore, u is unique; $\{f_a\}_{a\in A}$ are unique up to common, strictly increasing, transformation; and, for each $a \in A$, the probability measure $\pi(\cdot; a)$ is unique, satisfying $\pi(\theta; a) = 0$ if and only if θ is null given a.

As in Savage's (1954) model, the uniqueness of the probabilities in (10) is predicated on an arbitrary normalization of the functions u and $\{f_a\}_{a\in A}$.[8] Consequently, there is no guarantee that the prior probabilities correspond to the decision maker's prior beliefs.[9]

[8] The normalization consists of assigning b^{**} and b^* utilities as follows: $u(b^*(\theta), \theta) = 0$ for all $\theta \in \Theta$, and $\sum_{\theta\in\Theta} u(b^{**}(\theta), \theta)\, \pi(\theta, a) = 1$, $a \in A$. Then, for all $a \in A$, $f_a(1) = 1$ and $f_a(0) = 0$.

[9] To see this, fix a non-constant function $\gamma : \Theta \to \mathbb{R}_{++}$ and let $\Gamma(a) = \sum_{\theta\in\Theta} \gamma(\theta)\, \pi(\theta; a)$, $\hat{u}(b(\theta); \theta) = u(b(\theta); \theta)/\gamma(\theta)$, $\hat{\pi}(\theta; a) = \gamma(\theta)\, \pi(\theta; a)/\Gamma(a)$, and $\hat{f}_a = f_a \circ \Gamma(a)$. Then, by (10), the prior preference relation $\succcurlyeq$ is represented by

$$f_a\left(\Gamma(a) \sum_{\theta\in\Theta} \frac{u(b(\theta); \theta)}{\gamma(\theta)} \frac{\gamma(\theta)\, \pi(\theta; a)}{\Gamma(a)}\right) = \hat{f}_a\left(\sum_{\theta\in\Theta} \hat{u}(b(\theta); \theta)\, \hat{\pi}(\theta; a)\right). \tag{11}$$

But $\hat{\pi}(\theta; a) \neq \pi(\theta; a)$ for some a and θ. Therefore the prior subjective probabilities are not unique.

3 Bayesian Preferences and Subjective Probabilities

3.1 Bayesian decision makers

A Bayesian decision maker is characterized by a prior preference relation, a set of information-dependent posterior preferences, and the condition that the posterior preferences are obtained from the prior preference solely by updating his subjective probabilities using Bayes' rule (that is, leaving intact the other functions that figure in the representation). To model the behavior of Bayesian decision makers it is necessary to formally incorporate into the analytical framework new information.

Let X be a finite set of *observations.* For all $\theta \in \Theta$, $q(\cdot \mid \theta)$ is a likelihood function on X (that is, $q(\cdot \mid \theta) \geq 0$ and $\sum_{x \in X} q(x \mid \theta) = 1$). Elements of X are *observations* that might influence the decision maker's beliefs regarding the likelihoods of the alternative effects (experimental results, expert opinions, etc.). For every $x \in X$, let $\succcurlyeq^x$ be a (posterior) preference relation on $\mathbb{C}$ characterizing the decision maker's choice behavior after he has been informed of the observation x.

Consider a decision maker who, before choosing an action-bet pair, receives information pertinent to his assessment of the likely realization of alternative effects, conditional on his choice of action. Suppose that the decision maker updates his prior beliefs by the application of Bayes' rule. Formally,

Definition 2: A decision maker whose prior preference relation, $\succcurlyeq$ on $\mathbb{C}$, is represented by $f_a\left(\sum_{\theta\in\Theta} u\left(b\left(\theta\right);\theta\right)\pi\left(\theta;a\right)\right)$ is *Bayesian* if his posterior preference relations, $\{\succcurlyeq^x\}_{x\in X}$, are represented by $f_a\left(\sum_{\theta\in E} u\left(b\left(\theta\right);\theta\right)\pi\left(\theta\mid x,a\right)\right)$, where, for all $x\in X$, $\theta\in\Theta$ and $a\in A$,

$$\pi\left(\theta\mid x,a\right)=\frac{q\left(x\mid\theta\right)\pi\left(\theta;a\right)}{\sum_{\theta'\in\Theta} q\left(x\mid\theta'\right)\pi\left(\theta';a\right)}. \tag{12}$$

The use of objective likelihood functions in this context warrants some explanation. The effects are interpreted as conceivable unknown, parameters, whose likely realization depends on the actions. To grasp this consider the following example: A decision maker believes that the age of pregnant women affects the probability that their babies suffer form birth defects, and take this into consideration when she chooses the age of pregnancy. In this instance the age of the pregnancy corresponds to actions and birth defects and no defect correspond to effects. Alpha-Feto is a protein secreted by the fetal liver and excreted in the mother's bloodstream. Alpha-Feto protein test (AFP) is a blood test performed on pregnant women screening for neural tube defects such as a deformity of the spinal canal and the presence of Down syndrome. Let θ and θ' represent, respectively, the presence and absence of Down syndrome in a fetus. The likelihood functions correspond to the distribution of AFP in the blood sample conditional on whether or not the fetus dsplays the presence of Down syndrome.

3.2 The uniqueness of the priors

The main result asserts that if a decision maker is Bayesian in the sense of definition 2, then the prior, action-dependent, subjective probabilities representing his beliefs are identifiable. It requires, however, that the set of observations and associated likelihood functions be rich in the following sense: for every nonconstant function $\gamma : \Theta \to \mathbb{R}_{++}$ let $\Gamma_{\gamma}(a) := \sum_{\theta\in\Theta} \gamma(\theta) \pi(\theta \mid a)$ and $\Gamma_{\gamma}(a, x) := \sum_{\theta\in\Theta} \gamma(\theta) \pi(\theta \mid a, x)$, where $\pi(\theta \mid a)$ and $\pi(\theta \mid a, x)$ denote, respectively, the the prior probability of θ given a and the posterior probability of θ given a and x, then,

(R) For every nonconstant function $\gamma : \Theta \to \mathbb{R}_{++}$ there is an observation $x \in X$ and $a, a' \in A$ such that

$$\frac{\Gamma_{\gamma}(a)}{\Gamma_{\gamma}(a, x)} \geq 1 > \frac{\Gamma_{\gamma}(a')}{\Gamma_{\gamma}(a', x)}. \tag{13}$$

Theorem 1 *Suppose that there are at least two effects, R holds, and the decision maker's prior preference relation, $\succcurlyeq$ on $\mathbb{C}$, satisfies (A.0)–(A.3). If the decision maker is Bayesian, then the representation of his prior preferences admits a unique set of action-dependent probability measures.*

Proof. Consider a Bayesian decision maker whose prior preference relation $\succcurlyeq$ on $\mathbb{C}$ satisfies (A.0)–(A.3). Then, by Karni (2006), $\succcurlyeq$ has the representation in (10). Denote by $\succcurlyeq^{x}$ the decision maker's posterior preferences on $\mathbb{C}$ conditional on the observation $x \in X$. By definition, if the representation of $\succcurlyeq$ on $\mathbb{C}$ is given by

$$(a,b) \to f_a\left(\sum_{\theta\in\Theta} u\left(b\left(\theta\right);\theta\right)\pi\left(\theta;a\right)\right), \tag{14}$$

then for every $x \in X$, the posterior preference relation is represented by

$$(a,b) \mapsto f_a\left(\sum_{\theta\in\Theta} u\left(b\left(\theta\right);\theta\right)\pi\left(\theta\mid x,a\right)\right), \tag{15}$$

where $\pi\left(\theta\mid x,a\right)$ is given in (12).

Given a nonconstant function $\gamma:\Theta\to\mathbb{R}_{++}$ let $\Gamma_\gamma\left(a\right)=\sum_{\theta\in\Theta}\gamma\left(\theta\right)\pi\left(\theta;a\right)$.[10] Then the prior preferences $\succcurlyeq$ on $\mathbb{C}$ is represented by

$$(a,b) \to \hat{f}_a\left(\sum_{\theta\in\Theta} \hat{u}\left(b\left(\theta\right);\theta\right)\hat{\pi}\left(\theta;a\right)\right), \tag{16}$$

where $\hat{u}\left(b\left(\theta\right);\theta\right)=u\left(b\left(\theta\right);\theta\right)/\gamma\left(\theta\right)$, $\hat{\pi}\left(\theta;a\right)=\gamma\left(\theta\right)\pi\left(\theta;a\right)/\Gamma_\gamma\left(a\right)$, and $\hat{f}_a=f_a\circ\Gamma_\gamma\left(a\right)$. Moreover, by definition 2, for every $x \in X$, the corresponding posterior preference relation is represented by

$$(a,b) \mapsto \hat{f}_a\left(\sum_{\theta\in\Theta} \hat{u}\left(b\left(\theta\right);\theta\right)\hat{\pi}\left(\theta\mid x,a\right)\right), \tag{17}$$

where $\hat{\pi}\left(\theta\mid x,a\right)=q\left(x\mid\theta,a\right)\hat{\pi}\left(\theta;a\right)/\sum_{\theta'\in\Theta}q\left(x\mid\theta',a\right)\hat{\pi}\left(\theta';a\right)$. By definition,

$$\hat{\pi}\left(\theta\mid x,a\right)=\frac{q\left(x\mid\theta,a\right)\gamma\left(\theta\right)\pi\left(\theta;a\right)}{\sum_{\theta'\in\Theta}q\left(x\mid\theta',a\right)\gamma\left(\theta'\right)\pi\left(\theta';a\right)}. \tag{18}$$

Recall that $\Gamma_\gamma\left(a,x\right)=\sum_{\theta\in\Theta}\gamma\left(\theta\right)\pi\left(\theta\mid a,x\right)$. Hence, using (18),

$$\sum_{\theta\in\Theta}\hat{u}\left(b\left(\theta\right),\theta\right)\hat{\pi}\left(\theta\mid x,a\right)=\frac{1}{\Gamma_\gamma\left(a,x\right)}\sum_{\theta\in\Theta}u\left(b\left(\theta\right);\theta\right)\pi\left(\theta\mid x,a\right). \tag{19}$$

[10] Since γ is not constant, $\Gamma_\gamma\left(a\right)\neq 1$ for some $a \in A$.

Moreover, by definition,

$$\hat{f}_a\left(\sum_{\theta\in\Theta}\hat{u}\left(b\left(\theta\right),\theta\right)\hat{\pi}\left(\theta\mid x,a\right)\right)=f_a\left(\frac{\Gamma_\gamma\left(a\right)}{\Gamma_\gamma\left(a,x\right)}\sum_{\theta\in\Theta}u\left(b\left(\theta\right);\theta\right)\pi\left(\theta\mid x,a\right)\right). \qquad (20)$$

Next I show that there exists an observation $x\in X$ such that $\succcurlyeq^x\neq\hat{\succcurlyeq}^x$. Given the function γ above, take $a,a'\in A$ and $x\in X$ that satisfy R and $\hat{b}$, $\hat{b}'\in B$ be such that $u\left(\hat{b}\left(\theta\right),\theta\right)=\bar{u}$ and $u\left(\hat{b}'\left(\theta\right),\theta\right)=\bar{u}'$ for all $\theta\in\Theta$ and $f_a\left(\bar{u}\right)=f_{a'}\left(\bar{u}'\right)$. The existence of $\bar{b}$ and $\bar{b}'$ is an implication of (A.0) continuity of the uitlity functions and transitivity of $\succcurlyeq$. (To see this, take a constant-valuation bet $\bar{b}$, rearrange the effects so that $u\left(\bar{b}\left(1\right);1\right)\geq u\left(\bar{b}\left(2\right);2\right)\geq...\geq u\left(\bar{b}\left(n\right);n\right)$. If all the weak inequalities are eqaulities that $\bar{b}$ is a constant-utility bet and, by definition, $f_a\left(\bar{u}\right)=f_{a'}\left(\bar{u}'\right)$. Suppose that for some of the inequalities are strict. For all $\bar{b}\left(i\right)\in\arg\max\{u\left(\bar{b}\left(\theta\right),\theta\right)\mid\theta\in\Theta\}$ decrease the value of the payoff by the same amount and for all $\bar{b}\left(i\right)\in\arg\min\{u\left(\bar{b}\left(\theta\right),\theta\right)\mid\theta\in\Theta\}$ increase the value in such a way that $\sum_{\theta\in\Theta}u\left(b\left(\theta\right);\theta\right)\pi\left(\theta\mid a\right)$ remains constant. This is possible by continuity. Continue this process, until the payoff of the argmax set is equal to the second highest payoff or the payoff of the argmin set is equal to the second lowest payoff in $\{\bar{b}\left(\theta\right)\mid\theta\in\Theta\}$, whichever happens first. Repeat the process with the new armin and/or argmax sets. After a finite number of repetitions the payoffs, $\hat{b}\left(\theta\right)$, will be such that the utilities are the same. By construction $\left(a,\hat{b}\right)\sim\left(a,\bar{b}\right)$. Repeating the process with $\left(a',\bar{b}\right)$ to obtain $\hat{b}'$ such that $\left(a',\hat{b}'\right)\sim\left(a',\bar{b}\right)$). By definition $\left(a',\bar{b}\right)\sim\left(a,\bar{b}\right)$. Hence, by transitivity, $\left(a',\hat{b}'\right)\sim\left(a,\hat{b}\right)$, which implies $f_a\left(\bar{u}\right)=f_{a'}\left(\bar{u}'\right)$.

By the representation (15) $\left(a,\hat{b}\right)\sim^x\left(a',\hat{b}'\right)$, for all $x\in X$. Equation (20) and RC imply that

$$\begin{aligned}
\hat{f}_a\left(\sum_{\theta\in\Theta}\hat{u}\left(\hat{b}\left(\theta\right),\theta\right)\hat{\pi}\left(\theta \mid x,a\right)\right) &= f_a\left(\frac{\Gamma_\gamma\left(a\right)}{\Gamma_\gamma\left(a,x\right)}\bar{u}\right)\geq f_a\left(\bar{u}\right)= \qquad (21)\\
f_{a'}\left(\bar{u}'\right) &> f_{a'}\left(\frac{\Gamma_\gamma\left(a'\right)}{\Gamma_\gamma\left(a',x\right)}\bar{u}'\right)=\hat{f}_a\left(\sum_{\theta\in\Theta}\hat{u}\left(\hat{b}'\left(\theta\right),\theta\right)\hat{\pi}\left(\theta \mid x,a\right)\right) \qquad (22)
\end{aligned}$$

Thus $\left(a,\hat{b}\right)\hat{\succ}^x\left(a',\hat{b}'\right)$. Therefore $\succcurlyeq^x$ and $\hat{\succcurlyeq}^x$ cannot both be true. Consequently, there is a unique representation of the prior preference relations of a Bayesian decision maker that induces the correct posterior representations. The set of action-dependent probability measures in this representation is unique. ■

The proof of theorem 1 is a demonstration that updating the prior preference relation by applying Bayes' rule to a misspecified prior must, for some observations, induce posterior preference realtions that disagree with the decision maker's actual posterior preferences. To develop an intuitive understanding of this result, consider the case of additive representation, namely, $f_a\left(\sum_{\theta\in\Theta}u\left(b\left(\theta\right);\theta\right)\pi\left(\theta;a\right)\right)=\sum_{\theta\in\Theta}u\left(b\left(\theta\right);\theta\right)\pi\left(\theta;a\right)+v\left(a\right)$ or, equivalently, $\hat{f}_a\left(\sum_{\theta\in\Theta}\hat{u}\left(b\left(\theta\right);\theta\right)\hat{\pi}\left(\theta;a\right)\right)=\Gamma_\gamma\left(a\right)\sum_{\theta\in\Theta}\hat{u}\left(b\left(\theta\right);\theta\right)\hat{\pi}\left(\theta;a\right)+v\left(a\right),$ where $\hat{u}\left(b\left(\theta\right);\theta\right)=u\left(b\left(\theta\right);\theta\right)/\gamma\left(\theta\right),$ $\hat{\pi}\left(\theta;a\right)=\gamma\left(\theta\right)\pi\left(\theta;a\right)/\Gamma_\gamma\left(a\right),$ and $\Gamma_\gamma\left(a\right)=\sum_{\theta\in\Theta}\gamma\left(\theta\right)\pi\left(\theta;a\right).$

For every given a and *any* observation, x, the representation of the posterior preferences of a Bayesian decision maker is obtained by the application of Bayes' rule to the probability measure that figures in the prior representation. Hence the representations of the posterior preferences are: $\sum_{\theta\in\Theta}u\left(b\left(\theta\right);\theta\right)\pi\left(\theta \mid x,a\right)+v\left(a\right)$ or $\Gamma_\gamma\left(a\right)\sum_{\theta\in\Theta}\hat{u}\left(b\left(\theta\right);\theta\right)\hat{\pi}\left(\theta \mid x,a\right)+$

$v(a)$, where $\pi(\theta \mid x, a)$ is given in (12) and $\hat{\pi}(\theta \mid x, a)$ is obtained from $\hat{\pi}(\theta; a)$ by Bayes' formula. But, by definition,

$$\Gamma_\gamma(a) \sum_{\theta\in\Theta} \hat{u}(b(\theta); \theta) \hat{\pi}(\theta \mid x, a) + v(a) = \frac{\Gamma_\gamma(a)}{\Gamma_\gamma(a, x)} \sum_{\theta\in\Theta} u(b(\theta); \theta) \pi(\theta \mid x, a) + v(a).$$

Since $\Gamma_\gamma(a, x) > 0$, $\frac{\Gamma_\gamma(a)}{\Gamma_\gamma(a,x)} \sum_{\theta\in\Theta} u(b(\theta); \theta) \pi(\theta \mid x, a)$ is a positive increasing transformation of $\sum_{\theta\in\Theta} u(b(\theta); \theta) \pi(\theta \mid x, a)$. Hence for a given a, the induced posterior preferences are necessarily the same, regardless of representation of the prior preferences.

To obtain the uniqueness of the prior, it is necessary to consider the posterior preferences on bets *conditional on distinct actions*. Under the richness assumption the normalization associated with distinct actions makes for diverse multiplicative coefficients which imply that, for some observations,

$$\frac{\Gamma_\gamma(a)}{\Gamma_\gamma(a, x)} \sum_{\theta\in\Theta} u(b(\theta); \theta) \pi(\theta \mid x, a) + v(a) \geq \frac{\Gamma_\gamma(a')}{\Gamma_\gamma(a', x)} \sum_{\theta\in\Theta} u(b'(\theta); \theta) \pi(\theta \mid x, a') + v(a')$$

and

$$\sum_{\theta\in\Theta} u(b(\theta); \theta) \pi(\theta \mid x, a) + v(a) < \sum_{\theta\in\Theta} u(b'(\theta); \theta) \pi(\theta \mid x, a') + v(a').$$

At least one of these representations of the posterior preferences cannot be true. The same argument applies to the more general, nonadditive, representation (10).

4 Concluding Remarks

Insofar as Bayesian statistics is concerned, the main issue addressed by this work is the existence of identifiable prior representing the statistician's beliefs. As demonstrated in the introduction, subjective expected utility theories invoking Savage's analytical framework do not yield identifiable priors. In the theory presented here, the ability of a decision maker to influence the likely realization of effects by his choice of action, his manifested willingness to bet on the effects, and the changes in his betting behavior when new information becomes available yield sufficient additional data to identify the subjective probability representing the decision maker's beliefs.

A decision maker's beliefs are updated using likelihood functions that depict the relative frequency of potential observations conditional on the effects. To see how this model relates to the application of Bayes' rule in conventional statistical analysis, consider the following example. Let there be two sets of urns, K and K', and two types of urns. Urns of type 1 contain equal number of red and black balls, urns of type 2 contain twice as many red balls as black balls. We are interested in identifying the decision maker's prior probabilities regarding the type of a particular urn. Suppose that the decision maker is allowed to choose the set from which the urn is picked and simultaneously place bets on the type of the urn. Then the sets K and K' correspond to actions, the urns correspond to effects. Experiments consist sampling balls with replacement. Clearly, in this example, the likelihoods of the two colors conditional on the type of urn are "objective" in the sense of the relative frequencies while the prior and posteriors on the type of urn are subjective. If the statistician is Bayesian and his utility function is effect independent – a refutable hypothesis in this theory – then, by Theorem 1, the priors on the types corresponding to K and K' are identifiable.[11] In other

[11] In Karni (2006) I studied a special case of effect-independent preferences. In that case, the effect-

words, the model described above constitutes a choice-based foundations of the Bayesian prior.

The hybrid model presented here and the examples above entail subjective assessment of likelihoods of effects and objective (in the sense of relative frequencies) assessment of likelihoods of observations conditional on the effects. It is possible to extend the model to allow for subjective likelihood functions. In particular, the bets may be extended to include payoff contingent on observations. Formally, an *extended bet* is a function $\beta : \Theta \times X \to \mathbb{R}$ specifying monetary payoffs as a function of the observations and effects. Invoking the notational conventions introduced above, let $\beta_{-(\theta,x)}r$ be an extended bet whose (θ, x) coordinate is repalced by r. Define x to be a *null observation given* θ if θ is nonnull given a and $\left(a, \beta_{-(\theta,x)}r\right) \sim \left(a, \beta_{-(\theta,x)}r'\right)$ for all $r, r' \in \mathbb{R}$. An observation x is *certain given* θ if $\left(a, \beta_{-(\theta,x)}r\right) \succ \left(a, \beta_{-(\theta,x')}r'\right)$ for all $x' \neq x$ and $r, r' > 0$. For every θ if x is null given θ, set $q\left(x \mid \theta\right) = 0$ and if x is certain given θ, set $q\left(x \mid \theta\right) = 1$. These indicator functions are choice-based subjective likelihood functions the use of which renders the model purely subjective.

Subjective probability are intended to measure the degree of belief a decision maker has

dependent utility functions take the form $u\left(b\left(\theta\right), \theta\right) = \beta\left(\theta\right) u\left(b\left(\theta\right)\right) + \alpha\left(\theta\right), \beta\left(\theta\right) > 0$. The identifiability of the subjective probabilities in this model implies that the coefficients $\beta\left(\theta\right)$ and $\alpha\left(\theta\right)$ may be inferred from the decision maker's choice behavior. If the utility functions are effect independent (that is $\beta\left(\theta\right) = \beta$ and $\alpha\left(\theta\right) = \alpha$ for all θ) then it is natural to attribute to the model the usual statistical interpretation. Moreover, in this case constant-valuation bets are constant bets.

in the truth of events, that is, the degree of beliefs that the events obtain. This presumes that the attribute "truth of events" is a measurable cognitive object. Formally, a decision maker's beliefs is binary relation, $\trianglerighteq$ on $A \times 2^{\Theta}$, where $(a, E) \trianglerighteq (a', E')$ has the interpretation that the decision maker believes that E is more likely to obtain under the action a than E' is under a'. In the introduction to *Foundations of Measurement,* Krantz et. al. say: "When measuring some attribute of a class of objects or events, we associate numbers with the objects in such a way that the properties of the attribute are faithfully represented by the numerical properties." (Krantz et. al. (1971), p. 1). If the degree of belief a Bayesian decision maker has about the truth of an event is revealed by his choice behavior as hypothesized by Ramsey (1931), then the result obtained here implies that there exists an identifiable probability measure that faithfully represents the decision maker's beliefs.

References

[1] Drèze, J. H. 1987. Decision Theory with Moral Hazard and State-Dependent Preferences, in Drèze, J. H. Essays on Economic Decisions Under Uncertainty. (Cambridge University Press, Cambridge).

[2] Grossman, S. J., and O. D. Hart (1983) An Analysis of the Principal-Agent Problem, Econometrica 51, 7-45.

[3] Karni, E. 1993. A Definition of Subjective Probabilities with State-Dependent Preferences, Econometrica 61, 187-198.

[4] Karni, E. 1996. Probabilities and Beliefs. Journal of Risk and Uncertainty 13, 249-262.

[5] Karni, E. 2003. On the Representation of Beliefs by Probabilities. Journal of Risk and Uncertainty, 26, 17-38.

[6] Karni, E. 2006. Subjective Expected Utility Theory without States of the World. Journal of Mathematical Economics 42, 325 - 342.

[7] Karni, E. 2007. "A Foundations of Bayesian Theory." *Journal of Economic Theory,* 132, 167-188.

[8] Karni, E., D. Schmeidler. 1993. On the uniqueness of subjective probabilities. Economic Theory 3, 267-277.

[9] Krantz, D. E., R. D. Luce, P. Suppes, and A. Tversky. 1971. Foundations of Measurement. (Academic Press, New York and London).

[10] Nau, Robert F. (1995) Coherent Decision Analysis with Inseparable Probabilities and Utilities, Journal of Risk and Uncertainty 10, 71-91.

[11] Ramsey, Frank P. (1931) Truth and Probability, in The Foundations of Mathematics and Other Logical Essays. London: K. Paul, Trench, Truber and Co.

[12] Savage, L. J. 1954. The Foundations of Statistics. (John Wiley and Sons, New York).

[13] Schervish, M., J., Seidenfeldt, T. Kadane, J. B., 1990. State-Dependent Utilities. Journal of American Statistical Association 85, 840-847.

[14] Skiadas, C. 1997. Conditioning and Aggregation of Preferences. Econometrica 65, 347-367.

[15] Wakker, P. P. 1987. Subjective Probabilities for State-Dependent Continuous Utility. Mathematical Social Sciences 14, 289-298.

APPENDIX

A Moral Hazard Problem

The following moral hazard problem is formulated using the parametrized distribution approach. Let the agent be a risk-averse, Bayesian, expected utility–maximizing decision maker. Suppose that the principal is not only aware of this but that he also knows the agent's prior preferences. The principal-agent relationship is governed by a contract specifying the monetary payoff to the agent contingent on the outcome. If new pertinent information becomes available before the terms of the original contract are fulfilled, the contract is subject to renegotiation.

Even if it is state-independent, the agent's prior preference relation admits infinitely many representations, involving distinct probabilities and state-dependent utility functions (see Karni [1996]). Among these, presumably, is a unique probability measure that is the true representation of the agent's beliefs. I intend to show that, even if the principal knows the agent's prior preferences but ascribes to him the wrong probabilities (and consequently the wrong utilities), it is possible that once if new information becomes available, the two parties to the original contract may agree to replace it with a contract that induces the agent to select an action that is not in the principal's best interest.

For the sake of concreteness I consider the following example. At the end of a regatta, monetary prizes will be awarded to the boats that finish first and second. Suppose that there are three entrants, x, y, and z, and the prizes satisfy $m_1 > m_2 > m_3 = 0$, where m_i denotes the prize awarded to the boat in place i. According to the rules, if one of the boats withdraws from the race, it is automatically assigned third place; if two boats withdraw, the race is cancelled and no prizes are awarded. The relevant set of outcomes, Θ, consists of the six permutations of the order in which the boats cross the finish line.[12]

The boat owners recruit skippers and put them in charge of training their crews. Suppose that the owners cannot monitor the training and that the training entails costly effort on the part of skippers and crews. To motivate the sailors, an owner (principal) may offer his skipper (agent) an incentive contract.

In what follows I describe the initial incentive contract and examine the implications of renegotiating the contract of one of the remaining competitors - say, the owner of x - if one of the other entrants announces its withdrawal from the race. To simplify the analysis, I assume that the set of actions is a doubleton, with a' denoting training vigorously and a training lightly. Let $c : \{a, a'\} \to \mathbb{R}$ be a function depicting the disutility to the agent associated with the actions, and assume that $c(a') > c(a) = 0$. Let E_i denote the event that x finishes in position i (for example, $E_1 = \{(x, y, z), (x, z, y)\}$).

Recall that the skipper is assumed to be a risk-averse, Bayesian, expected utility - maximizer, and that *the boat owner knows the skipper's prior preferences, and that the skipper*

[12] The outcomes of the race correspond to effects. As this example illustrates, the distinguishing characteristic of effects is that, unlike states, their likely realization may be influenced by the decision maker.

is a Bayesian. Suppose that the boat owner ascribes to the skipper his own probabilities regarding the likely outcome, conditional on the level of training. This is the common prior assumption, often invoked in the parametrized distribution formulation of principal-agent problems.[13] Under these assumptions, if the principal wants to induce the agent to train vigorously, his problem may be stated as follows:

Program 1: Choose a contract $\mathbf{w} := (w(E_1), w(E_2), w(E_3)) \in \mathbb{R}^3$ so as to maximize

$$\sum_{i=1}^{3} u(m_i - w(E_i))\, \pi(E_i; a') \tag{23}$$

subject to the individual rationality constraint

$$\sum_{i=1}^{3} v(w(E_i))\, \pi(E_i; a') - c(a') \geq \bar{v} \tag{24}$$

and the incentive compatibility constraint

$$\sum_{i=1}^{3} v(w(E_i))\, \pi(E_i; a') - c(a') \geq \sum_{i=1}^{3} v(w(E_i))\, \pi(E_i; a) - c(a), \tag{25}$$

where u and $\{\pi(E_i, j) \mid i \in \{1,2,3\},\ j \in \{a, a'\}\}$ denote the utility functions and the conditional (on the actions) subjective probabilities of the principal, and v and c denote, respectively, the utility of money and the disutility associated with the actions ascribed by the principal to the agent. By assumption, $\pi(\cdot, \cdot)$ also denote the conditional probabilities ascribed by the principal to the agent.

By assumption, *the principal employs a representation of the agent's prior preferences*

[13] This entails a loss of generality, as the skipper's prior preferences must permit this type of manipulations of the representation. I indicate below where this assumption comes into play.

consistent with the agent's choice behavior. This depiction of the principal's problem agrees with the parametrized distribution formulation in the literature on agency theory.

Representation of the Agent's Preferences

It is conceivable that, independently of his monetary reward, the agent cares about winning but that the principal is unaware of this.[14] Moreover, the principal cannot detect the agent's sentiments about winning from the skipper's choice behavior. Suppose that the agent's true utility function is given by $V(w, i, j) = \Gamma(j) v(w) / \gamma(i)$, where $0 < \gamma(1) < \gamma(2) < \gamma(3)$, and $\Gamma(j) = \sum_{i=1}^{3} \gamma(i) \pi(E_i; j)$, $j \in \{a, a'\}$. Note that the presence of a non-additive component, $\Gamma(j)$, in the agent's utility of action is also undetectable by observing his choice behavior.[15] Consistency of the agent's prior preferences with their representation in program 1

requires that his beliefs be represented by the probabilities

$$\pi^A(E_i; j) = \frac{\gamma(i) \pi(E_i; j)}{\Gamma(j)}, \quad j \in \{a, a'\}, \ i = 1, 2, 3. \tag{26}$$

Hence the agent's true objective function is given by:

$$\Gamma(j) \sum_{i=1}^{3} \gamma^{-1}(i) v(w(E_i)) \pi^A(E_i; j) - c(j), \quad j \in \{a, a'\}. \tag{27}$$

Program 1 may then be restated a follows:

[14] This is, of course, a simplifying assumption. In general, it is sufficient that the prinicpal not know *how much* the agent cares about winning.

[15] See Grossman and Hart (1983) for an example of principal-agent analysis in which the utility of action has a multiplicative factor, as above.

PROGRAM 1': Choose a contract, $\mathbf{w} \in \mathbb{R}^3$, so as to maximize

$$\sum_{i=1}^{3} u\left(m_i - w\left(E_i\right)\right) \pi\left(E_i; a'\right) \tag{28}$$

subject to the individual rationality constraint

$$\sum_{i=1}^{3} V\left(w\left(E_i\right), i, a'\right) \pi^A\left(E_i; a'\right) - c\left(a'\right) \geq \bar{v} \tag{29}$$

and the incentive compatibility constraint

$$\sum_{i=1}^{3} V\left(w\left(E_i\right), i, a'\right) \pi^A\left(E_i; a'\right) - c\left(a'\right) \geq \sum_{i=1}^{3} V\left(w\left(E_i\right), i, a\right) \pi^A\left(E_i; a\right). \tag{30}$$

Clearly, the solution, $\mathbf{w}^*$, to the two programs is the same. Thus insofar as the principal is concerned, as long as the contract $\mathbf{w}^*$ is in effect, the misspecification of the agent's utilities and probabilities is of no consequence.

More generally, this discussion shows that, from a substantive point of view, when possible, ascribing to the agent the principal's own beliefs is immaterial, provided that the utility functions are adjusted so that the agent's preferences are accurately represented. This implies that the "common prior" assumption is harmless if the agent's preferences are represented correctly and *no new information becomes available before the terms of the contract are fulfilled.* I revisit the "common prior" assumption at the end of this section.

New Information and Recontracting

Suppose that after having signed a contract, the owner and skipper of boat x learn that one of the entrants – say, z – decides to withdraw from the race. This decision eliminates the possibility that one of the two remaining boats finishes third. Equivalently, the event consisting of all the outcomes in which z finishes first or second becomes null. Let $E_3(z)$ denote the event consisting of outcomes in which z finishes third. Since both the principal and the agent are Bayesian, the principal's posterior probabilities are $\pi(E_i; j \mid E_3(z)) = \pi(E_3(z); j \mid E_i)\,\pi(E_i; j)/\pi(E_3(z); j)$, and $\pi(E_3; j \mid E_3(z)) = 0$, $j \in \{a, a'\}$. Moreover, being aware that the agent is Bayesian, the principal updates the agent's probabilities in the same way. However – and here's the rub – the agent's *true* posterior probabilities are $\pi^A(E_3; j \mid E_3(z)) = \pi^A(E_3(z); j \mid E_i)\,\pi^A(E_i; j)/\pi^A(E_3(z); j)$, and $\pi^A(E_3; j \mid E_3(z)) = 0$, $j \in \{a, a'\}$. To simplify the notation, henceforth I denote the posteriors $\pi(E_i; j \mid E_3(z))$ and $\pi^A(E_i; j \mid E_3(z))$ by $\hat{\pi}(E_i, j)$ and $\hat{\pi}^A(E_i, j)$, respectively.

Consider next the issue of recontracting. The withdrawal of z from the race raises two questions. First, are there perceived benefits to recontracting? Second, if the original contract is replaced, is it necessarily true that the replacement contract induces the agent to implement the action desired by the principal?

In principle, the original and the replacement contracts may be negotiated simultaneously or sequentially. In other words, given that the announcement that one of the competitors has pulled out of the race is made public before the beginning of training, the contract governing the principal-agent relationship in this event may be negotiated before or after the announcement. However, except for the cost of contracting, there is no difference between the simultaneous and sequential negotiations. If the contracts are negotiated simultaneously, the reservation utility of the agent must be the same under both contracts. If the contracts are negotiated sequentially, the first contract can be written to include a clause that makes it null and void if one of the competitors withdraws from the race. The second contract is negotiated under the same individual rationality and incentive compatibility constraints and is, therefore, no different from the contract covering the same contingency that was negotiated simultaneously. In the presence of contracting costs, the sequential negotiation is preferable on two counts. First, if none of the competitors pulls out of the race, there is no need for a second contract, which saves the contracting cost. Second, if one of the competitors does pull out, it is better to incur the cost of recontracting later than earlier.

In what follows, I pursue the scenario of sequential contracting and analyze the case in which the principal wants the agent to implement the rigorous training program. From the principal's point of view, the subsequent contract, $\mathbf{w}^{**} = \left(w^{**}\left(E_1\right), w^{**}\left(E_2\right)\right)$, is the solution to the following program:

Program 2: Choose $\mathbf{w} \in \mathbb{R}^2$ so as to maximize

$$\sum_{i=1}^{2} u\left(m_i - w\left(E_i\right)\right) \hat{\pi}\left(E_i; a'\right) \tag{31}$$

subject to the individual rationality constraint

$$\sum_{i=1}^{2} v\left(w\left(E_i\right)\right) \hat{\pi}\left(E_i; a'\right) - c\left(a'\right) \geq \bar{v} \tag{32}$$

and the incentive compatibility constraint

$$\sum_{i=1}^{2} v\left(w\left(E_i\right)\right) \hat{\pi}\left(E_i; a'\right) - c\left(a'\right) \geq \sum_{i=1}^{2} v\left(w\left(E_i\right)\right) \hat{\pi}\left(E_i; a\right) - c\left(a\right). \tag{33}$$

Recalling that $c(a) = 0$, it is easily verifiable that the solution, $\mathbf{w}^{**}$, to program 2, is given by the solution to the following equations:

$$v(w(E_1)) = \bar{v} + \frac{[1 - \hat{\pi}(E_1; a)]\, c(a')}{\hat{\pi}(E_1; a') - \hat{\pi}(E_1; a)} \text{ and } v(w(E_2)) = \bar{v} - \frac{\hat{\pi}(E_1; a)\, c(a')}{\hat{\pi}(E_1; a') - \hat{\pi}(E_1; a)}. \tag{34}$$

Consider next the problem as viewed by the agent. The agent should agree to $\mathbf{w}^{**}$ if, under the corresponding best action, it meets his reservation utility. Formally, $\mathbf{w}^{**}$ is acceptable to the agent if

$$\max_{j \in \{a, a'\}} \sum_{i=1}^{2} V(w^{**}(E_i), i, j)\, \hat{\pi}^A(E_i; j) - c(j) \geq \bar{v}. \tag{35}$$

The second question - namely, whether it is necessarily true that $\mathbf{w}^{**}$ induces the agent to implement the action desired by the principal - may be restated as follows: Does the incentive compatibility constraint

$$\sum_{i=1}^{2} V(w^{**}(E_i), i, a')\, \hat{\pi}^A(E_i; a') - c(a') \geq \sum_{i=1}^{2} V(w^{**}(E_i), i, a)\, \hat{\pi}^A(E_i; a) \tag{36}$$

hold?

I am interested in showing that it is possible that inequality (36) fails to hold. In this instance, if z pulls out of the race, the original contract becomes null and void and the two parties to the contract sign a new contract, $\mathbf{w}^{**}$. However, under the new contract, contrary to the expectations and the interests of the principal, the agent implements the light training program. In other words, *because he failed to ascribe to the agent the correct prior probabilities and utilities, the principal misrepresents the agent's posterior preference* *relation and, as a result, fails to provide the agent with the incentives that would have induced him to act in the principal's best interest.*

A Numerical Example

In order to simplify the calculations, assume that the principal is risk neutral. Let the agent's utility function be $V(w,i,j) = \Gamma(j)\sqrt{w}/\gamma(i)$, $i \in \{1,2,3\}$, $j \in \{a,a'\}$. Suppose that the principal perceives the problem facing him as program 1 above, with the prior probabilities as follows:

$$\begin{array}{lcccccc} \textit{Outcomes} & \begin{pmatrix} x \\ y \\ z \end{pmatrix} & \begin{pmatrix} x \\ z \\ y \end{pmatrix} & \begin{pmatrix} y \\ x \\ z \end{pmatrix} & \begin{pmatrix} z \\ x \\ y \end{pmatrix} & \begin{pmatrix} y \\ z \\ x \end{pmatrix} & \begin{pmatrix} z \\ y \\ x \end{pmatrix} \\ \pi(\cdot;a) & \frac{1.5}{12} & \frac{1.5}{12} & \frac{2}{12} & \frac{2}{12} & \frac{2.5}{12} & \frac{2.5}{12} \\ \pi(\cdot;a') & \frac{4}{12} & \frac{4}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} & \frac{1}{12} \end{array} \tag{37}$$

where, in describing the outcomes, the boats are ranked, from top to bottom, according to their positions. Recall that the principal is the owner of boat x. Hence the principal's prior probabilities of the relevant events are:

$$\begin{array}{lll} \pi(E_1;a) = \frac{3}{12} & \pi(E_2;a) = \frac{4}{12} & \pi(E_3;a) = \frac{5}{12} \\ \pi(E_1;a') = \frac{8}{12} & \pi(E_2;a') = \frac{2}{12} & \pi(E_3;a') = \frac{2}{12} \end{array} \tag{38}$$

Maintain the assumption that $c(a) = 0$, and suppose that the agent's disutility associated with the rigorous training is $c(a') = 0.1$ and his prior reservation utility level is $\bar{v} = 11.8333$. Consider the contract

$$\mathbf{w}^* = \left(w^*(E_1) = 144.26\ , w^*(E_2) = 137.82\ , w^*(E_3) = 139.65\right). \tag{39}$$

Insofar as the principal is concerned, this contract satisfies the agent's perceived individual rationality constraint

$$\frac{8}{12}\sqrt{144.26} + \frac{2}{12}\sqrt{137.82} + \frac{2}{12}\sqrt{139.65} - 0.1 = 11.8333 \tag{40}$$

and the incentive compatibility constraint[16]

$$11.8333 = \frac{3}{12}\sqrt{144.26} + \frac{4}{12}\sqrt{137.82} + \frac{5}{12}\sqrt{139.65}. \tag{41}$$

Moreover, the principal would rather the agent implement a' under the contract $\mathbf{w}^*$ than a under the fixed-pay contract $\bar{\mathbf{w}} = 11.8333^2$. To see this, note that the value of the principal's objective function under $(\mathbf{w}^*, a')$ is

$$\frac{8}{12}(1000 - 144.26) + \frac{2}{12}(500 - 137.82) - \frac{2}{12}139.65 = 607.58, \tag{42}$$

far exceeding the value of his objective function under the fixed contract and a, which is

$$\frac{3}{12}1000 + \frac{4}{12}500 - 11.833^2 = 276.65. \tag{43}$$

[16] The fact that in this example the incentive compatibility constraint is binding does not affect the generality of the conclusion below.

Next consider the situation as seen by the agent. Suppose that the agent's true prior beliefs are represented by the action dependent probabilities as follows:

$$\begin{array}{ccccccc} Outcomes & \begin{pmatrix} x \\ y \\ z \end{pmatrix} & \begin{pmatrix} x \\ z \\ y \end{pmatrix} & \begin{pmatrix} y \\ x \\ z \end{pmatrix} & \begin{pmatrix} z \\ x \\ y \end{pmatrix} & \begin{pmatrix} y \\ z \\ x \end{pmatrix} & \begin{pmatrix} z \\ y \\ x \end{pmatrix} \\ \pi^A(\cdot; a) & \frac{32{,}847}{1{,}000{,}000} & \frac{32{,}847}{1{,}000{,}000} & \frac{175{,}180}{1{,}000{,}000} & \frac{175{,}180}{1{,}000{,}000} & \frac{291{,}970}{1{,}000{,}000} & \frac{291{,}970}{1{,}000{,}000} \\ \pi^A(\cdot; a') & \frac{15}{100} & \frac{15}{100} & \frac{15}{100} & \frac{15}{100} & \frac{20}{100} & \frac{20}{100} \end{array} \tag{44}$$

This means that there are positive numbers $\gamma(i)$, $i = 1, 2, 3$, satisfying $\gamma(1) < \gamma(2) < \gamma(3)$, and $\Gamma(j) = \sum_{i=1}^{3} \gamma(i) \pi(E_i; j)$, $j \in \{a, a'\}$, such that $\pi^A(E_i; j) = \gamma(i) \pi(E_i; j) / \Gamma(j)$. The solution to $\gamma(i)$, $i = 1, 2, 3$, and $\Gamma(j)$, $j \in \{a, a'\}$, is $\gamma(1) = 6.0336, \gamma(2) = 24.134$, $\gamma(3) = 32.179$, $\Gamma(a') = 13.408$, and $\Gamma(a) = 22.961$.[17]

[17] The calculation of $\{\gamma(i) \mid i = 1, 2, 3\}$ and $\Gamma(j)$, $j \in (a, a')$ is based on the following equations:

$$\frac{\frac{8}{12}\gamma(1)}{\frac{8}{12}\gamma(1)+\frac{2}{12}\gamma(2)+\frac{2}{12}\gamma(3)} = 0.3$$

$$\frac{\frac{2}{12}\gamma(2)}{\frac{8}{12}\gamma(1)+\frac{2}{12}\gamma(2)+\frac{2}{12}\gamma(3)} = 0.3$$

$$\frac{\frac{2}{12}\gamma(3)}{\frac{8}{12}\gamma(1)+\frac{2}{12}\gamma(2)+\frac{2}{12}\gamma(3)} = 0.4$$

the solution to which is

$$\gamma(1) = 6.0336, \gamma(2) = 24.134, \gamma(3) = 32.179$$

Thus

$$\Gamma(a) = \frac{3}{12}6.0336 + \frac{4}{12}24.134 + \frac{5}{12}32.179 = 22.961$$

and

$$\Gamma(a') = \frac{8}{12}6.0336 + \frac{2}{12}24.134 + \frac{2}{12}32.179 = 13.408.$$

Hence his probabilities of the relevant events are[18]

$$\begin{array}{ccc} \pi^A(E_1;a) = 0.065694 & \pi^A(E_2;a) = 0.35036 & \pi^A(E_3;a) = 0.58394 \\ \pi^A(E_1;a') = 0.3 & \pi^A(E_2;a') = 0.3 & \pi^A(E_3;a') = 0.4 \end{array} \tag{45}$$

Because $\mathbf{w}^*$ in (39) is the solution of program 1, it satisfies the individual rationality and incentive compatibility constraints. Thus the agent accepts the contract $\mathbf{w}^*$ and, under its terms, implements the rigorous training program, as expected by the principal.

Consider next the effect of z pulling out of the race. The principal's posterior probabilities, which are also the posterior probabilities he ascribes to the agent, are obtained by the application of Bayes' rule. They are given by:

$$\begin{array}{cc} \hat{\pi}(E_1;a) = \frac{15}{35} & \hat{\pi}(E_2;a) = \frac{20}{35} \\ \hat{\pi}(E_1;a') = \frac{4}{5} & \hat{\pi}(E_2;a') = \frac{1}{5} \end{array} \tag{46}$$

The principal considers offering the agent the contract, $\mathbf{w}^{**} = (w^{**}(E_1), w^{**}(E_2))$, that would motivate him to implement a'. This contract is obtained by the solution to the equations (34) and is given by $w^{**}(E_1) = 143.69$, and $w^{**}(E_2) = 137.31$.[19]

[18] The conditional probabilities $\{\pi^A(E_i;a) \mid i = 1,2,3\}$ are the solution to the equations

$$\frac{\frac{3}{12}6.0336}{\frac{3}{12}6.0336+\frac{4}{12}24.134+\frac{5}{12}32.179} = \pi^A(E_1;a)$$
$$\frac{\frac{4}{12}24.134}{\frac{3}{12}6.0336+\frac{4}{12}24.134+\frac{5}{12}32.179} = \pi^A(E_2;a)$$
$$\frac{\frac{5}{12}32.179}{\frac{3}{12}6.0336+\frac{4}{12}24.134+\frac{5}{12}32.179} = \pi^A(E_3;a)$$

[19] Specifically, $\mathbf{w}^{**}$ is given by the equations

The principal's payoff under $(\mathbf{w}^{**}, a')$, is[20]

$$\frac{4}{5}(1000 - 143.69) + \frac{1}{5}(500 - 137.31) = 757.6 \tag{47}$$

His payoff under $(\bar{\mathbf{w}}, a)$, is

$$\frac{15}{35}1000 + \frac{20}{35}500 - 11.8333^2 = 574.26. \tag{48}$$

Clearly, the principal is better off with $(\mathbf{w}^{**}, a')$ than with $(\bar{\mathbf{w}}, a)$. That is, the principal is better off offering to replace contract $\mathbf{w}^{*}$ by contract $\mathbf{w}^{**}$, expecting the agent to implement the rigorous training program.

Next I show that the agent accepts $\mathbf{w}^{**}$ and, contrary to the anticipations of the principal, chooses action a. Note that the agent's posterior beliefs are represented by

$$\begin{array}{cc} \hat{\pi}^A(E_1; a) = \frac{32{,}847}{32{,}847+175{,}180} = 0.1579 & \hat{\pi}^A(E_2; a) = \frac{175{,}180}{32{,}847+175{,}180} = 0.8421 \\ \hat{\pi}^A(E_1; a') = 0.5 & \hat{\pi}^A(E_2; a') = 0.5 \end{array} \tag{49}$$

If the agent accepts $\mathbf{w}^{**}$ and implements a, his posterior expected utility is

$$22.961\left(\frac{0.1579\sqrt{143.69}}{6.0336} + \frac{0.8421\sqrt{137.31}}{24.134}\right) = 16.591, \tag{50}$$

$\sqrt{w^{**}(E_1)} = 11.8333 + \frac{20\times 0.1}{13}$, and $\sqrt{w^{**}(E_2)} = 11.8333 - \frac{15\times 0.1}{13}$.

[20] The principal believes that if the agent acepts $\mathbf{w}^{**}$, it will induce him to select the action a'.

which exceeds his reservation utility. Thus the agent accepts $\mathbf{w}^{**}$. The agent's payoff under $(\mathbf{w}^{**}, a')$ is

$$13.408 \left(\frac{0.5\sqrt{143.69}}{6.0336} + \frac{0.5\sqrt{137.31}}{24.134} \right) - 0.1 = 16.474. \tag{51}$$

Comparing the agent's payoffs corresponding to $\mathbf{w}^{**}$ under the two actions, it is clear that $\mathbf{w}^{**}$ is not incentive compatible. Hence *against the principal's best interest (and wishes), the agent implements the light training program.*

Conclusion and Explanation

This example illustrates the potential consequences of the principal's failure to ascribe to the agent his true prior probabilities and utilities, even if the representation of the agent's prior preferences is accurate. It undermines the use of the common prior assumption, which pervades agency theory. It shows that if the agent's preferences are outcome independent, the principal knows the agent's prior preferences, and no new information becomes available, the common prior assumption is harmless and convenient. However, in situations in which recontracting under new information is possible, the use of the common prior assumption is no longer harmless or automatically justified. In particular, if the posterior contracts are to implement the second-best solutions, the agent's true prior needs to be identified, and it may or may not agree with the principal's.

The result in this example reflects the fact that the agent believes that his choice of the more costly action shifts relatively large probability mass from the worst outcome, E_3, to the best outcome, E_1, but little from E_2 to E_1. Hence the agent readily responds to incentives when E_3 is feasible. However, once E_3 has been eliminated, the agent's responsiveness diminishes. Because he ascribed to the agent the wrong probabilities, the principal fails to appreciate this lack of responsiveness and offers the agent the "wrong" replacement contract.