Why Does it Matter that Beliefs and Valuations be Correctly Represented?[*]

Abstract

This paper contains an analysis of a simple principal-agent problem illustrating possible problems that may arise when the prinicpal ascribes to the agent subjective probabilities and utilities that are implied by the subjective expected utility model but do not represent the agent's beliefs and valuations. In particular, it is possible that an incentive contract designed by the principal induces the agent to choose an action that is not in the principal's best interest.

Keywords: subjective probability, moral hazard, state-dependent preferences.

JEL classifications: D81, D82

Simon Grant
Department of Economics
Rice University
email: sgrant@rice.edu

Edi Karni
Department of Economics
Johns Hopkins University
email: karni@jhunix.hcf.jhu.edu

---

1   Introduction

Subjective expected utility theory is founded on the tacit notion that choice among alternative courses of action (acts) is governed by two separate cognitive processes: the assessment of the likelihood of various events, or the formation of beliefs, and the valuation of the consequences associated with those events. Moreover, beliefs are supposed to be coherent enough to allow their representation by a (subjective) probability measure, and the valuation of the consequences sufficiently structured to permit their representation by numerical utilities. Individual preferences on acts are represented by the expected values of the utilities of the consequences of these acts with respect to the subjective probability measure.

Choice-theoretic models of subjective expected utility, including Savage (1954) and Anscombe and Aumann (1963), derive the subjective probabilities and utilities from individuals' preference relations on the set of acts. These, and all similarly conceived, models give rise to equivalent representations of preferences each involving a utility function and a corresponding subjective probability measure. To determine a unique subjective probability the choice-theoretic models invoke the convention, not implied by the axioms, that the utility function is state independent. Put differently, the axiomatic structures of the various choice-theoretic subjective expected utility models require that the preference relations on acts be state independent (for example, Savage's postulates P3 and P4, and Anscombe and Aumann state-independence axiom) but that does not imply that the utility function must be state-independent. In fact, state-independent preferences, that is, state-independent risk attitudes, only require that the utility function representing the valuation of the consequences in different states be positive affine transformations of one another. Thus the normalization of the utility functions to make them the same across states has no compelling choice-theoretic foundation. Moreover, the subjective probabilities are the normalized multiplicative coefficients of these utility functions. Hence these subjective probabilities are inherently arbitrary (see Schervish, Seidenfeld, and Kadane [1990]; Karni and Schmeidler [1993]; and Karni [1996], [2003]). In particular, *it is possible that a decision maker's preference relation on acts satisfy the axioms of subjective expected utility theory and yet be ascribed probabilities that do not represent his beliefs and utilities that do not represent his valuations.*

To remedy this problem and obtain a definition of subjective probability that quantifies the decision-makers' beliefs it is necessary to extend the choice space. One possible extension, due to Karni and Schmeidler (1980), calls for the introduction of a second preference relation over hypothetical lotteries on the set of state-consequence pairs. The new preference relation is linked axiomatically to the preference relation on horse/roulette-lotteries acts in the framework of Anscombe and Aumann (1963). The original intent of Karni and Schmeidler was to model subjective expected utility theory with state-dependent preferences, however, Karni and Mongin (2000) recently noted that the probabilities thus obtained are, in fact, the unique correct representation of decision makers' beliefs. The model of Karni and Schmeidler, as well as the more general expected utility model explored in Karni (2003) and the nonexpected utility theory developed in Grant and Karni (2003), in all of which the subjective probabilities represent of decision-makers' beliefs, rely on the use of objective probabilities on the set of states as a primitive concept. More recently, Karni (2003) developed an axiomatic subjective expected utility model with preferences defined on conditional acts (or, alternatively, preference over actions that delimit the events that might obtain) that leads to a definition of subjective probabilities representing decision makers' beliefs, and utilities that represent their valuations. Like Savage (1954), Karni's theory does not involve the use of objective probabilities as a primitive concept but, unlike Savage, it accommodates state-dependent preferences and applies to situations involving moral hazard.

These developments raise the question: other than for philosophical reasons, why is it important to represent decision makers' beliefs and valuations correctly? Karni (1996) argued that such a representation is desirable since it renders the decision-makers' observed choice behavior and their verbal exchange of information consistent. Karni (2003) showed, in the context of a simple principal-agent problem, that if the agent is an expected utility maximizing Bayesian decision maker and principal ascribes to the agent probabilities, that misrepresent the agent's beliefs, and designs an incentive-contract based on these probabilities, the principal runs the risk of inducing the agent to choose an action that is not in the principal's best interest.

Our purpose, in this paper, is to explore this issue further. In particular, we intend to examine the role of ascribing the agent the correct utility function. To do this we show that if the principal

ascribes incorrect utilities and/or subjective probabilities to the agent, he may fail to induce the agent to act in a way that serves the best interest of the principal. In other words, we show that a contract designed, on the basis of ascribed probabilities and utilities implied by choice-theoretic subjective expected utility model, to motivate the agent to choose one action motivates him, instead, to choose another action that is less desirable for the principal.

2 The Envious Agent Problem

In classical economic theory self-interest seeking behavior is portrayed strictly as a quest to improve the individual's material well-being. This narrow view of human nature has recently been challenged and the possibility of incorporating emotions into the theory of choice is explored (see, for example, a survey by Elster [1998] and discussions by Loewenstein [2000], Romer [2000]). The interest in broadening the psychological basis underlaying the conduct of economic agents is due, in part, to experimental evidence indicating a tendency of individuals to cooperate in situations in which maximization of material self-interest alone would imply non-cooperative behavior (see Camerer [1997], Berg, Dickhaut, and McCabe [1995]). Against this backdrop, we consider next a principal-agent relation that may be influenced by envy. Specifically, we analyze a principal-agent problem in which the agent's preferences incorporate envy, and yet are representable by a subjective expected utility functional. We show that the failure of the principal to detect the presence of envy results in a contract, based on the principal's ascribed utilities and probabilities, that motivates the agent to act in a way that is not in the principal's best interest.

2.1 An advertising campaign

Consider the following principal-agent problem. A producer (the principal) engages an advertising agency to promote an event (for example, a rock concert). The revenue is a random variable that depends on the state of nature which, in this instance, represents the state of demand and on the advertising campaign. Specifically, suppose that there are three states of nature $S = \{H, M, L\}$, where H signifies high demand, M signifies moderate demand, and L signifies low demand.

The agent must choose between actively promoting the event, choosing the action a^1 and incurring the cost (disutility) $v(a^1) > 0$, and doing nothing, namely, choosing the costless default action a^0, (that is, $v\left(a^0\right) = 0$). Assume that if he does not actively promote the event, then if the demand is high the producer will sell the concert-hall capacity and attain the high level of revenue, r^H, if the demand is moderate he will sell half of the concert-hall capacity and attain a revenue of r^M, and if the demand is low he will sell only 15 percent of the concert-hall capacity and attain low revenue level, r^L. If the agent undertakes an active advertising campaign to reach a wide audience he can boost the demand to the point of ensuring himself of selling at least half of the concert-hall capacity. In other words, he can prevent the situation in which only 15 percent of the capacity is sold, and will either sell half the concert-hall capacity or the entire concert-hall capacity. Assume that the nature of the advertising campaign (effort and cost invested to reach the potential audience) is private information of the agent.

To model the situation described above let $\mathbf{B} = \{a^1, a^0\}$ denote the set of feasible actions. The effects of the alternative advertising campaigns are expressed by the mapping $F : \mathbf{B} \rightarrow \mathcal{E}$, where $\mathcal{E}$ is the set of events (that is, subsets of the set S). Thus the active advertising campaign a^1 corresponds to the event $F\left(a^1\right) = \{H, M\}$ and a^0 corresponds to the universal event $F\left(a^0\right) = \{H, M, L\}$. Note that, once the agent chooses an action, say $a \in \mathbf{B}$, the elements of the set $F\left(a\right)$ correspond to Savage's definition of the state of the world (nature), namely, "a description of the world, leaving no relevant aspect undescribed," (Savage 1954, p. 9) since, given a, the state of demand alone determines the revenue, which is the only relevant aspect of nature. The fact that the set of states depends on the action means that our framework is that of preferences on conditional acts (see Luce and Krantz [1971], Fishburn [1973], and Karni [2003]).

Assume that both the principal and the agent are expected utility-maximizing Bayesian decision makers whose preferences are state independent. This terminology merits some elaboration. First, a subjective expected utility maximizing decision maker is Bayesian if he updates his prior subjective probabilities using Bayes rule. While the choice-theoretic subjective expected utility model does not imply this particular updating rule, it is, nevertheless, consistent with it. Subjective expected utility models in which Bayesian updating is implied require the extension of

4

the analytical framework. For example, Ghirartato (2002) uses conditional preferences on acts and Karni (2003) uses preferences on conditional acts to obtain subjective expected utility representations of Bayesian decision makers' preferences. Second, as noted above, the framework that we use is that of preferences on conditional acts. This means that for every $a \in \mathbf{B}$, the set of states is $F(a)$, and the corresponding (conditional) acts are functions from $F(a)$ to the set of consequences. A decision maker is a subjective expected utility maximizer if his preferences on conditional acts are representable by a subjective expected utility functional. The axiomatic foundations of subjective expected utility theory of Bayesian decision making, which is the theory used here, is developed in Karni (2003).

Suppose that the principal and the agent are Bayesian decision makers and that their beliefs regarding the likely realization of the states conditional on the acts are the same. Specifically, conditional on the agent's choosing the action a^0 these beliefs are represented by the probability distribution $\left(\pi^H, \pi^M, \pi^L\right)$. However, the principal and the agent are unaware of this uniformity of their beliefs. Assume that the principal is risk neutral and her utility function is state independent (that is, the principal's utility function is the identity function) and that the agent is risk averse and that his state-dependent valuations of the payoff $w \in [0, 1]$ (where 1 corresponds to $1 million) are depicted by state-dependent utility functions $u^H(w) = \beta^H u(w) + \alpha^H$, $u^M(w) = \beta^M u(w) + \alpha^M$, and $u^L(w) = \beta^L u(w) + \alpha^L$, where u is strictly concave, continuously differentiable, real-valued function. Without loss of generality let $u(1) = 1$, $u(0) = 0$, $\sum_{s \in S} \pi^s \beta^s = 1$ and $\sum_{s \in S} \pi^s \alpha^s = 0$.

To depict the agent's choice behavior we adopt the framework of Anscombe-Aumann (1963) in which acts (unconditional and conditional) are functions from the relevant set of states to the set of simple monetary lotteries and the agent's preference relation is a state-independent and continuous weak order, $\succsim$. In this framework, if the principal could verify that the agent employed the action a^0 and observe his choice behavior he would ascribe to the agent subjective probabilities, $p = \{p^H, p^M, p^L\}$ and a state-independent utility function, $\hat{u}(w)$, implied by the choice-theoretic subjective expected utility model. For example, the principal can set $\hat{u}(0) := 0$ and $\hat{u}(1) := 1$. Then, for each $E \in \mathcal{E}$, p^E is the unique 'probability' in $[0, 1]$ which renders the agent indifferent between the constant act $p^E \delta_1 + \left(1 - p^E\right) \delta_0$ and the act whose payoff is δ_1 in the

5

event E and is δ_0 in the complementary event $S - E$, where δ_w denotes the degenerate monetary lottery that assigns to the payoff w the unit probability mass. Formally:

$$p^E \delta_1 + \left(1 - p^E\right) \delta_0 \sim \begin{bmatrix} \delta_1 & s \in E \\ \\ \delta_0 & s \notin E \end{bmatrix}. \tag{1}$$

Hence, for each $s \in S$, we obtain

$$p^s \left[\pi^H \left(\beta^H + \alpha^H\right) + \pi^M \left(\beta^M + \alpha^M\right) + \pi^L \left(\beta^L + \alpha^L\right)\right] = \pi^s \beta^s + \sum_{s \in S} \pi^s \alpha^s. \tag{2}$$

But $\sum_{s \in S} \pi^s \beta^s = 1$ and $\sum_{s \in S} \pi^s \alpha^s = 0$, hence, if no active advertising campaign is undertaken by the agent, the principal ascribes to the agent the belief that state s obtains with probability $p^s = \pi^s \beta^s$. Moreover, the principal ascribes to the agent the probability $p\left(\{H, M\}\right) = p^H + p^M$ for the event $\{H, M\}$.

Similarly, for each $w \in [0, 1]$, $\hat{u}\left(w\right)$ is set equal to the unique probability u_w for which

$$u_w \delta_1 + \left(1 - u_w\right) \delta_0 \sim \delta_w. \tag{3}$$

So we obtain

$$u_w \left[\pi^H \left(\beta^H + \alpha^H\right) + \pi^M \left(\beta^M + \alpha^M\right) + \pi^L \left(\beta^L + \alpha^L\right)\right] \tag{4}$$
$$= \pi^H \left(\beta^H u\left(w\right) + \alpha^H\right) + \pi^M \left(\beta^M u\left(w\right) + \alpha^M\right) + \pi^L \left(\beta^L u\left(w\right) + \alpha^L\right),$$

that is, $\hat{u}\left(w\right) = u\left(w\right)$.

Furthermore, for $s \in F\left(a^1\right) = \{H, M\}$, the principal ascribes to the agent the conditional belief $p^{s|\{H,M\}}$, which is the unique 'probability' in $[0, 1]$ that satisfies

$$p^{s|\{H,M\}} \begin{bmatrix} \delta_1 & s' \in \{H, M\} \\ \\ \delta_0 & s' = L \end{bmatrix} + \left(1 - p^{s|\{H,M\}}\right) \delta_0 \sim \begin{bmatrix} \delta_1 & s' = s \\ \\ \delta_0 & s' \in \{H, M\} - \{s\} \\ \\ \delta_0 & s = L \end{bmatrix}. \tag{5}$$

That is, for $s \in F\left(a^1\right)$ we obtain

$$p^{s|\{H,M\}} \left(p^H + p^M\right) = p^s \tag{6}$$

and so

$$p^{s|\{M,H\}} = \frac{p^s}{p^H + p^M}. \tag{7}$$

6

Finally, if the disutility to the agent of undertaking an active advertising campaign is equal to $v(a^1) > v(a^0) = 0$, the principal will ascribe to the agent for the contract $\mathbf{w} = (w_H, w_M, w_L)$ that pays w_s in state $s \in \{H, M, L\}$ the subjective expected utility

$$\frac{p^H}{p^H + p^M} u(w_H) + \frac{p^M}{p^H + p^M} u(w_M) - v(a^1). \tag{8}$$

2.2 Principal-agent problems

Let the state-contingent revenue $\mathbf{r} = (r^H, r^M, r^L)$ satisfy $r^H > r^M > r^L$. A contract, $\mathbf{w}$, is a point in $\mathbb{R}^3_+$ representing the agent's state contingent pay. We assume that contracts that entail net negative wealth for the agent are not feasible.[1] Then, given the principal's subjective probabilities, π_P, and his perception of the agent's subjective probabilities, p_A, and utility, u, if the principal wants to induce the agent to undertake the advertising campaign, his problem may be stated as follows:

Choose $\mathbf{w} \in \mathbb{R}^3_+$ so as to maximize

$$\sum_{s \in S} \pi_P\left(s \mid F\left(a^1\right)\right)(r^s - w_s) \tag{9}$$

subject to the participation constraint:

$$\sum_{s \in S} p_A\left(s \mid F\left(a^1\right)\right) u(w_s) - v\left(a^1\right) \geq v_0, \tag{10}$$

and the incentive compatibility constraint:

$$\sum_{s \in S} p_A\left(s \mid F\left(a^1\right)\right) u(w_s) - v\left(a^1\right) \geq \sum_{s \in S} p_A\left(s \mid F\left(a^0\right)\right) u(w_s), \tag{11}$$

where v_0 is the outside option available to the agent in case he rejects the contract. Let $v_0 = 0$. The principal will offer the contract $\mathbf{w}$ if

$$\sum_{s \in S} \pi_P\left(s \mid F\left(a^1\right)\right)(r^s - w_s) \geq \sum_{s \in S} \pi_P\left(s \mid F\left(a^0\right)\right)(r^s)$$

Note that $p_A\left(s \mid F\left(a\right)\right) = p^s / p\left(F\left(a\right)\right)$ for all $a \in \mathbf{B}$ and $s \in S$.

---

[1] For example, the agent may not have any resources of her own and be liquidity constrained.

The agent's problem may be stated as follows: Given the optimal contract $\mathbf{w}^*$ choose $a \in \mathbf{B}$ so as to maximize

$$U\left(a; \mathbf{w}^*\right) = \sum_{s \in S} \pi_A\left(s \mid F\left(a\right)\right)\left(\beta^s u\left(w_s^*\right) + \alpha^s\right) - v(a) \tag{12}$$

and implement the optimal action a^* if $U\left(a^*; \mathbf{w}^*\right) \geq v_0$. Otherwise reject the contract.

It is obvious that the principal's perception of the agent's motives is different from the agent's true motives. To illustrate the potential pitfalls of misconstrued assignment of utilities and probabilities to the agent, we analyze two specific cases in which the agent is envious of the principal. We then state a more general result, the Envious Agent Theorem.

2.3 Case I

Let the agent's envy affect the marginal utility of his income. Specifically, suppose that increases in the principal's revenue reduces the agent's marginal utility of his income uniformly. To capture this trait of the agent's attitudes and at the same time preserve the preference structure, we let $\alpha^H = \alpha^M = \alpha^L = 0$ and suppose that $0 < \beta^H < \beta^M < \beta^L$. The assumption $\beta^L > \beta^M > \beta^H$ is given the interpretation that, for any given level of w, the agent's utility and his marginal utility of income are higher the lower is the principal's income. This malevolent attitude cannot be detected by observing the agent's choice behavior over unconditional acts. Notice that, combined with $\sum_{s \in S} \pi^s \beta^s = 1$, this implies that $\beta^L > 1$.

From the viewpoint of the principal the expected revenue if he does not induce the agent to undertake an active advertising campaign is $\pi^H r^H + \pi^M r^M + \pi^L r^L$.

Supposing that the agent is Bayesian and that the prior probabilities (that is, the probabilities conditional on a^0) are p^s, $s \in \{H, M, L\}$, the principal perceives the incentive compatibility constraint to induce the agent to undertake an advertising campaign to be:

$$\frac{p^H}{p^H + p^M} u\left(w_H\right) + \frac{p^M}{p^H + p^M} u\left(w_M\right) - v\left(a^1\right) \geq p^H u\left(w_H\right) + p^M u\left(w_M\right) + p^L u\left(w_L\right) > 0. \tag{13}$$

But expression (13) implies that the participation constraint is not binding. Hence, the first order

conditions are:

$$-\frac{\pi^s}{\pi^H + \pi^M} + \lambda p^s u'(w_s^*) \frac{p^L}{p^H + p^M} = 0, \ s \in \{H, M\}, \tag{14}$$

where w_s^* denote the optimal value of w_s. Hence

$$\frac{\pi^H}{\pi^M} = \frac{p^H u'(w_H^*)}{p^M u'(w_M^*)}. \tag{15}$$

But $\beta^M > \beta^H$ implies that $p^H/p^M < \pi^H/\pi^M$. Thus $u'(w_M^*) < u'(w_H^*)$ and the optimal contract requires that $w_M^* > w_H^* > w_L^* = 0$. Despite having the same beliefs regarding the likely realizations of the states and the fact that the principal being risk neutral while the agent is risk averse, the state-dependence of the agent's utility requires that he bears some risk under the optimal contract. Note also that under the optimal contract the agent's payoff is not monotonic in the revenue.

Equation (15) implies that

$$\beta^M u'(w_M^*) = \beta^H u'(w_H^*). \tag{16}$$

Making the first inequality in expression (13) an equality, we see that a binding incentive compatibility constraint as perceived by the principal entails:

$$v(a^1) = \frac{p^L}{1 - p^L} \left[p^H u(w_H^*) + p^M u(w_M^*) \right]. \tag{17}$$

since our normalization entails $u(0) = 0$.

Let us assume that the disutility to the agent of implementing the action a^1 is sufficiently small and r^H, r^M sufficiently large so that the principal perceives it to be in her best interest to induce the agent to undertake the advertising campaign. That is,

$$\frac{\pi^H \left(r^H - w_H^* \right) + \pi^M \left(r^M - w_M^* \right)}{\pi^H + \pi^M} > \pi^H r^H + \pi^M r^M + \pi^L r^L.$$

Consider next the agent's choices among the action-act pairs $(a^1, \mathbf{w}^*)$ and $(a^0, \mathbf{w}^*)$. According to the agent's beliefs, the probabilities of the event $F(a^1)$ is $\pi^A(\{H, M\}) = \pi^H + \pi^M$. Let $U(a, \mathbf{w}^*)$ be the agent's subjective expected utility corresponding to the action-act pair $(a, \mathbf{w}^*)$, $a \in \{a^1, a^0\}$. Then,

$$U(a^1; \mathbf{w}^*) = \frac{1}{1 - \pi^L} \left[\pi^H \beta^H u(w_H^*) + \pi^M \beta^M u(w_M^*) \right] - v(a^1)$$

$$U(a^0, \mathbf{w}^*) = \pi^H \beta^H u(w_H^*) + \pi^M \beta^M u(w_M^*)$$

And so

$$U\left(a^1; \mathbf{w}^*\right) - U\left(a^0, \mathbf{w}^*\right) = \frac{\pi^L}{1 - \pi^L}\left[p^H u\left(w_H^*\right) + p^M u\left(w_M^*\right)\right] - v\left(a^1\right)$$

Substituting for $v\left(a^1\right)$ from equation (17) into the previous expression yields

$$U\left(a^1; \mathbf{w}^*\right) - U\left(a^0, \mathbf{w}^*\right) = \left[\frac{\pi^L}{1 - \pi^L} - \frac{p^L}{1 - p^L}\right]\left[p^H u\left(w_H^*\right) + p^M u\left(w_M^*\right)\right] < 0.$$

The last inequality follows from the fact that $\beta^L > 1$ implies the belief p^L perceived by the principal is greater than the actual belief π^L held by the agent.

That is, given the contract $\mathbf{w}^*$ which he accepts (since the participation constraint is not binding) the agent chooses *not* to incur the expense and undertake an active advertising campaign, contrary to the wishes of the principal. *Because she misconstrued the agent's subjective probabilities and utilities, the principal designed a contract that induced the agent to choose an action that is not in the principal's best interest.*

2.4 Case II

Envy may manifest itself by affecting the level of the agent's utility without, at the same time, affecting his marginal utilities. In this case the principal ascribes to the agent probabilities that accurately reflects his beliefs. Yet, by misunderstanding the agent's motives the principal still fails to induce him to choose the desirable action. To analyze this situation we let $\alpha^H < \alpha^M < \alpha^L$, and $\beta^H = \beta^M = \beta^L = 1$. In this case the agent's envy is captured by values of the additive constants. Specifically, $\alpha^H < \alpha^M < \alpha^L$ is interpreted to mean that, for any given level of w, the agent's utility is higher the lower is the principal's income. Notice that, combined with $\sum_{s \in S} \pi^s \alpha^s = 0$, this implies that $\alpha^L > 0$.

As before, the principal's problem is to design a contract $\mathbf{w}^*$ that will implement a^1. This time the probabilities that the principal ascribes to the agent agrees with the agent's own probabilities. Furthermore since the agent's marginal utility of consumption is state-independent we can immediately conclude from the fact that the agent is risk averse and the principal is risk neutral, the optimal contract requires that $w_L^* = 0$ and $w_M^* = w_H^* = w^*$.

From the principal's perspective the incentive compatibility constraint is,

$$u\left(w^*\right) - v\left(a^1\right) \geq \left(\pi^H + \pi^M\right)u\left(w^*\right). \tag{18}$$

10

The incentive compatibility constraint is binding and the solution is the w^* for which expression (18) holds with equality. That is,

$$\pi^L u(w^*) = v\left(a^1\right) \tag{19}$$

As in case 1, assume that $v\left(a^1\right)$ is sufficiently small, so that the principal's perceived expected utility under $\left(a^1; \mathbf{w}^*\right)$ is greater than the expected revenue from having no advertising campaign, that is,

$$\frac{\pi^H r^H + \pi^M r^M}{\pi^H + \pi^M} - w^* > \pi^H r^H + \pi^M r^M + \pi^L r^L.$$

Next consider the problem from the viewpoint of the agent. For all $v\left(a^1\right) > 0$

$$
\begin{aligned}
U\left(a^0; \mathbf{w}^*\right) &= \pi^H \left(u\left(w^*\right) + \alpha^H\right) + \pi^M \left(u\left(w^*\right) + \alpha^M\right) + \pi^L \left(u\left(0\right) + \alpha^L\right) \\
&= \left(\pi^H + \pi^M\right) u\left(w^*\right) = \frac{\left(\pi^H + \pi^M\right)}{\pi_L} v\left(a^1\right) \\
&> \frac{\left(\pi^H + \pi^M\right)}{\pi_L} v\left(a^1\right) - \frac{\pi^L \alpha^L}{\pi^H + \pi^M} = \left(\pi^H + \pi^M\right) u\left(w^*\right) - \frac{\pi^L \alpha^L}{\pi^H + \pi^M} \\
&= \frac{\pi^H}{\pi^H + \pi^M} \left[u\left(w^*\right) + \alpha^H\right] + \frac{\pi^M}{\pi^H + \pi^M} \left[u\left(w^*\right) + \alpha^M\right] - v\left(a^1\right) \\
&= U\left(a^1; \mathbf{w}^*\right).
\end{aligned}
$$

Moreover, $U\left(a^0; \mathbf{w}^*\right) > 0$ implies that the participation constraint is satisfied. Hence, the agent chooses a^0, which again is the action that was not in the principal's best interest.

2.5 The envious agent theorem

The cases above illustrate a more general result. Let the elements of the state space, $S = \{1, ..., n\}$ be ordered so that the payoff of the principal is monotonic increasing in the states, that is $r^{s+1} \geq r^s$ for all $s \in \{1, ..., n-1\}$. An agent is said to be envious of the principal if his state-dependent utility function, $\beta^s u\left(\cdot\right) + \alpha^s$ displays the following properties: $\alpha^{s+1} \leq \alpha^s$ and $\beta^{s+1} \leq \beta^s$, where either $\alpha^{s+1} < \alpha^s$ for all $s = 1, \ldots n-1$ or $\beta^{s+1} < \beta^s$ for all $s = 1, \ldots n-1$. Let $a^0, a^1, ..., a^{n-1}$ be actions such that $F\left(a^j\right) = S - \{1, ..., j\}$, where $\{1, ..., 0\} = \emptyset$ and $v\left(a^{j+1}\right) > v\left(a^j\right)$ for all $j = 0, ..., n$. The principal-agent problem is said to have a nontrivial solution if the principal wants to induce the agent to take an action other than the default action a^0. We summarize the above discussion in the following theorem:

Theorem 1 *Suppose that the principal is risk neutral and that the agent is risk averse and envious. If the agent's preferences and actions are private information and the principal-agent problem has a nontrivial solution then the contract perceived to be optimal by the principal will induce the agent to take an action that is suboptimal from the viewpoint of the principal.*

Proof. Suppose that the agent is Bayesian and that the prior probabilities (that is, the probabilities conditional on a^0) ascribed to him by the principal are p^s, $s \in \{1, ..., n\}$. Let a^j be the optimal action from the point of view of the principal. Since the problem has a nontrivial solution, $n - 1 \geq j > 0$. Let $\mathbf{w}^*$ denote the optimal contract. Denote $\Pi(k) = \sum_{s=k+1}^{n} \pi^s$ and $P(k) = \sum_{s=k+1}^{n} p^s$, $k = 0, ..., n-1$. The principal perceives the incentive compatibility constraints necessary to induce the agent to undertake the action a^j to be:

$$\sum_{s=j+1}^{n} \frac{p^s}{P(j)} u(w_s^*) - v(a^j) \geq \sum_{s=k+1}^{n} \frac{p^s}{P(k)} u(w_s^*) - v(a^k), \; k = 0, ..., n-1. \tag{20}$$

Claim 1: The participation constraint is not binding.

To prove Claim 1 consider the agent's choices among the action-act pairs $(a^j, \mathbf{w}^*)$ and $(a^0, \mathbf{w}^*)$. Then the incentive compatibility constraints imply

$$\sum_{s=j+1}^{n} \left(\frac{1}{P(j)} - \frac{1}{P(0)} \right) p^s u(w_s^*) \geq v(a^j) - v(a^0) > 0. \tag{21}$$

But $P(0) = 1$ and $v(a^0,) = 0$. Hence

$$\sum_{s=j+1}^{n} p^s u(w_s^*) = \frac{P(j) v(a^j)}{1 - P(j)} > 0. \tag{22}$$

implies that the participation constraint is not binding. $\square$

Claim 1 implies that $w_s^* = 0$ for $s = 1, ..., j - 1$.

Claim 2: For some $k < j$ the incentive compatibility constraint is binding.

To prove Claim 2 note that, for $k < j$, $P(k) > P(j)$ and $v(a^k) < v(a^j)$. Suppose, by way of negation, that for all $k < j$ the incentive compatibility constraints are not binding, then since $w_s^* = 0$ for all $s = 1, ..., j$ we get:

$$\sum_{s=j+1}^{n} \frac{p^s}{P(j)} u(w_s^*) - v(a^j) > \sum_{s=j+1}^{n} \frac{p^s}{P(k)} u(w_s^*) - v(a^k), \quad k = 1, ..., j - 1. \tag{23}$$

Because the participation constraint is not binding, there is another contract $\bar{\mathbf{w}}$ such that $\bar{w}_s = 0$ for all $s = 1, ..., j-1$, and $\bar{w}_s < w_s^*$ for all $s = j, ..., n$ that induces the agent to implement a^j. This contradicts the optimality of $\mathbf{w}^*$. $\square$

Consider next the agent's choice among the action-act pairs $\left(a^j, \mathbf{w}^*\right)$ and $\left(a^k, \mathbf{w}^*\right)$, where $k < j$ is an index for which the incentive compatibility constraint is binding, that is,

$$\sum_{s=j+1}^{n} \frac{p^s}{P(j)} u\left(w_s^*\right) - v\left(a^j\right) = \sum_{s=j+1}^{n} \frac{p^s}{P(k)} u\left(w_s^*\right) - v\left(a^k\right). \tag{24}$$

From the viewpoint of the agent:

$$
\begin{aligned}
U\left(a^k, \mathbf{w}^*\right) &= \sum_{s=k+1}^{n} \frac{\pi^s}{\Pi(k)}\left[\beta^s u\left(w_s^*\right) + \alpha^s\right] - v\left(a^k\right) \\
&= \sum_{s=j+1}^{n} \frac{\pi^s}{\Pi(k)}\beta^s u\left(w_s^*\right) + \sum_{s=k+1}^{n} \frac{\pi^s}{\Pi(k)}\alpha^s - v\left(a^k\right) \\
&= \frac{P(k)}{\Pi(k)} \sum_{s=j+1}^{n} \frac{\pi^s}{P(k)}\beta^s u\left(w_s^*\right) + \sum_{s=k+1}^{n} \frac{\pi^s}{\Pi(k)}\alpha^s - v\left(a^k\right) \\
&> \frac{P(j)}{\Pi(j)} \sum_{s=j+1}^{n} \frac{\pi^s}{P(j)}\beta^s u\left(w_s^*\right) + \sum_{s=j+1}^{n} \frac{\pi^s}{\Pi(j)}\alpha^s - v\left(a^j\right) \\
&= \sum_{s=j+1}^{n} \frac{\pi^s}{\Pi(j)}\beta^s u\left(w_s^*\right) + \sum_{s=j+1}^{n} \frac{\pi^s}{\Pi(j)}\alpha^s - v\left(a^j\right) \\
&= \sum_{s=j+1}^{n} \frac{\pi^s}{\Pi(j)}\left[\beta^s u\left(w_s^*\right) + \alpha^s\right] - v\left(a^j\right) \\
&= U\left(a^j, \mathbf{w}^*\right)
\end{aligned}
$$

(25)

where the inequality follows from equation (24) and the following facts:

$$\sum_{s=k+1}^{n} \frac{\pi^s}{\Pi(k)}\alpha^s \geq \sum_{s=j+1}^{n} \frac{\pi^s}{\Pi(j)}\alpha^s$$

with strict inequality if $\alpha^s < \alpha^{s+1}$ for every $s = 1, \ldots, n-1$, and

$$\frac{P(k)}{P(j)} = 1 + \frac{\sum_{s=k+1}^{j-1} p^s}{P(j)} \geq 1 + \frac{\sum_{s=k+1}^{j-1} \pi_s}{\Pi(j)} = \frac{\Pi(k)}{\Pi(j)} \tag{26}$$

with strict inequality if $\beta^s < \beta^{s+1}$ for every $s = 1, \ldots, n-1$. To grasp the last inequality note that

$$\frac{\sum_{s=k+1}^{j-1} p^s}{P(j)} = \frac{\sum_{s=k+1}^{j-1} \beta^s \pi^s}{\sum_{s=j}^{n} \beta^s \pi^s} \geq \frac{\beta^{j-1} \sum_{s=k+1}^{j-1} \pi^s}{\beta^j \sum_{s=j}^{n} \pi^s} \geq \frac{\sum_{s=k+1}^{j-1} \pi^s}{\sum_{s=j}^{n} \pi^s}. \tag{27}$$

13

The implication of inequality (25) is that, contrary to the interest of the principal, the agent will not implement a^j. ∎

The Envious Agent Theorem only stated that, insofar as the principal is concerned, the action that will be implemented is suboptimal. In fact, it is possible to show that the action that will be implemented, say a^m, must be less costly to the agent than the action desired by the principal (that is, $m < j$). For example, if actions represent levels of effort and a^k implies smaller level of effort than a^{k+1}, then the optimal contract induces a level of effort smaller that desired by the principal. To show this we first characterize the nature of the optimal contract as perceived by the principal:

Proposition 2 *Suppose that the principal is risk neutral and that the agent is risk averse and envious. If the agent's preferences and actions are private information and the principal-agent problem has a nontrivial solution then the contract perceived to be optimal by the principal requires that $w_i^* = 0$ for $i = 1, ..., j$, and $w_i^* \leq w_{i+1}^*$ for $i = j+1, ..., n-1$, with $w_i^* > w_{i+1}^*$ if $\beta^i > \beta^{i+1}$.*

Proof. Observe that the first order conditions of the principal agent problem are:

$$-\frac{\pi^s}{\Pi(j)} + p^s u'(w_s^*) \sum_{k=1}^{s-1} \frac{\lambda_k}{P(k)} + \mu \frac{p^s}{P(j)} u'(w_s^*) = 0, \ s \in \{j, ..., n\}, \tag{28}$$

where λ_k are the Lagrange multipliers corresponding to the incentive compatibility constraints and μ is the Lagrange multiplier corresponding to the participation constraint. Since the latter constraint is not binding, $\mu = 0$. Equations (28) imply that, for all $s, t \in \{j, ..., n\}$,

$$\frac{\pi^s}{\pi^t} = \frac{p^s u'(w_s^*) \sum_{k=1}^{s-1} \frac{\lambda_k}{P(k)}}{p^t u'(w_t^*) \sum_{k=1}^{t-1} \frac{\lambda_k}{P(k)}}. \tag{29}$$

where w_s^* denote the optimal values of w_s. But $\beta^{i+1} \leq \beta^i$ implies that $p^i/p^{i+1} \geq \pi^i/\pi^{i+1}$. Moreover $\sum_{k=1}^{i} \frac{\lambda_i}{P(i)} \leq \sum_{k=1}^{i+1} \frac{\lambda_i}{P(i)}$. Thus $u'(w_i^*) \leq u'(w_{i+1}^*)$ and the optimal contract requires that $w_i^* \geq w_{i+1}^*$ for $i = j+1, ..., n-1$ and $w_i^* = 0$ for $i = 1, ..., j$. with $w_i^* > w_{i+1}^*$ if $\beta^i > \beta^{i+1}$. ∎

To show that the action implemented by the optimal contract. $\mathbf{w}^*$, is less costly to the agent than the action desired by the principal it is sufficient to establish that, for every $k > j$ the incentive compatibility constraint under the optimal contract is not binding.

Proposition 3 *Suppose that the principal is risk neutral and that the agent is risk averse and envious. If the agent's preferences and actions are private information and the principal-agent problem has a nontrivial solution, a^j, then under the contract perceived to be optimal by the principal, for all $k > j$*

$$\sum_{s=j+1}^{n} \frac{\pi^s}{\Pi(j)} \left[\beta^s u\left(w_s^*\right) + \alpha^s\right] - v\left(a^j\right) > \sum_{s=k+1}^{n} \frac{\pi^s}{\Pi(k)} \left[\beta^s u\left(w_s^*\right) + \alpha^s\right] - v\left(a^k\right).$$

Proof. From Proposition 1, we have $u\left(w_i^*\right) = 0$ for $i = 1, ..., j$, and $u\left(w_i^*\right) \geq u\left(w_{i+1}^*\right)$ for $i = j+1, ..., n-1$, with $u\left(w_i^*\right) > u\left(w_{i+1}^*\right)$ if $\beta^i > \beta^{i+1}$. Hence $\beta^i u\left(w_i^*\right) + \alpha^i \geq \beta^{i+1} u\left(w_{i+1}^*\right) + \alpha^{i+1}$ for all $i = j+1, ..., n-1$. Thus, by choosing an action a^k, where $k > j$, the agent incurs an increased cost since $v\left(a^k\right) > v\left(a^j\right)$ and only succeeds in eliminating states where his utility is maximal. ∎

3   Concluding Remarks

The analysis in this paper illustrates and underscores the possible pitfalls of employing subjective expected utility theory to the analysis of principal-agent problems. The source of difficulty is that the agent's preferences may admit alternative equivalent representations involving distinct subjective probabilities and state-dependent utility functions. If one's only concern is with individual decisions and is willing to assume state-independent preferences, then nothing essential is lost by imposing the convention that the utility functions are state-independent and defining subjective probabilities consistent with this convention. Decision makers' beliefs, namely, a binary relation on the set of events depicting the notion of "more likely to obtain," are *defined* by the probabilities. In other words, if the only application of the theory is to individual decision making then it is not necessary to separate utility and true probability, since only the product of the two matters. Our analysis shows that this is no longer the case if the model is to be applied to the richer context of the principal-agent theory. We analyzed a simple example but the reader will recognize that the issue pervades the entire principal-agent literature.

Strictly speaking subjective expected utility theory a la Savage (1954) does not apply to situations in which the decision maker is capable, by taking an appropriate action, to prevent the realization of certain events. To grasp this point, consider again the agent in the example above.

15

Let $\mathbf{x} = (x_L, x_M, x_H)$, where x_s denotes the monetary payoff in state s. Then $\mathbf{x}$ is an act in the sense of Savage (1954). The agent's expected utility corresponding to the action-act pairs $(a^0; \mathbf{x})$ and $(a^1; \mathbf{x})$ are given, respectively by:

$$U\left(a^0; \mathbf{x}\right) = \left[\pi^H \left(\beta^H u\left(x_H\right) + \alpha^H\right) + \pi^M \left(\beta^M u\left(x_M\right) + \alpha^M\right)\right] + \pi^L \left(\beta^L u\left(x_L\right) + \alpha^L\right) \quad (30)$$

and

$$U\left(a^1; \mathbf{x}\right) = \frac{\pi^H}{\pi^H + \pi^M} \left(\beta^H u\left(x_H\right) + \alpha^H\right) + \frac{\pi^M}{\pi^H + \pi^M} \left(\beta^M u\left(x_M\right) + \alpha^M\right) - v\left(a^1.\right) \quad (31)$$

Now, given $v\left(a^1\right) > 0$, let $\bar{x} \in (0,1)$ be the solution of the equation:[2]

$$U\left(a^0; (\bar{x}, \bar{x}, 0)\right) = U\left(a^1; (\bar{x}, \bar{x}, 0)\right). \quad (32)$$

Let $x'' > \bar{x} > x'$ and $\varepsilon > 0$. Then the act (x', x', ε) is strictly preferred by the agent over the act $(x', x', 0)$ yet, for x'' sufficiently large and ε sufficiently small, the act $(x'', x'', 0)$ is indifferent to the act (x'', x'', ε), contradicting Savage's Sure Thing Principle. Put differently, even though the agent is subjective expected utility maximizing Bayesian decision maker, his choice behavior is inconsistent with the core axiom of the theory. To describe this agent's choice behavior it is necessary to apply the subjective expected utility model piecewise, namely, to describe his behavior by the utility function $U\left(a^0; \mathbf{x}\right)$ on the subset of acts $\mathbf{X}^0 = \{\mathbf{x} \in [0,1]^3 \mid U\left(a^0; \mathbf{x}\right) < U\left(a^1; \mathbf{x}\right)\}$ and by the utility function $U\left(a^1; \mathbf{x}\right)$ on the complementary set $\mathbf{X}^1 = [0,1]^3 - \mathbf{X}^0$. The agent behaves as if his preferences vary over the choice set while, in fact, the preferences are invariant. The misperception is due to the misspecification of the choice set, that is, the disregard of the fact that the alternatives the agents must choose from are action-act pairs and not acts alone. Karni (2003) provides an axiomatic subjective expected utility model in which the choice set consists of action-act pairs that is applicable to the analysis of the principal-agent problem described above.

Our analysis imposes the restriction that contracts stipulating a payment by the agent to the principal in some events are not enforceable. It is well known that if such payments could be enforced it would be possible to penalize the agent to coerce him to avoid taking certain actions that may be detected, ex post. In our example, if a large penalty could be imposed in case the

---

[2] Such a solution exist if $\left(-\frac{\pi^H}{\pi^H + \pi^M}\right)\left(\beta^H + \alpha^H\right) + \left(-\frac{\pi^M}{\pi^H + \pi^M}\right)\left(\beta^M + \alpha^M\right) > \alpha^L + v\left(a^1\right) / \pi^L > 0$

revenue r^L is realized it would be possible to force the agent to avoid the action a^0 for fear of being detected and penalized after the fact. In the literature on principal-agent problem this issue is dealt with by assuming that all the conditional probability distributions *on outcomes* have the same support (see Salanié [1997]). To justify our approach we note that, as a matter of fact, it may be impossible, in some situations, to enforce the required penalty. On the theoretical level we note that the traditional analyses of the principal-agent problem (for example, Holmstrom [1979], Shavell [1979]) suppress the explicit consideration of the states of nature and focus instead on the conditional probabilities of the random variables representing the payoff to the principal. Our analytical framework can accommodate the traditional formulation if we assume that there are many states and that the outcome r^L corresponds to an event in the state space and that the agent may, by choosing an appropriate action, nullify a subevent, so that the conditional probability of r^L changes but is not zero. Then the probability distributions over outcomes corresponding to different actions all have exactly the same support even though, in the state-space the essential support of the posterior is a proper subset of that of the prior.

References

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

Berg, Joyce, John Dickhaut and Kevin McCabe (1995) "Trust, Reciprocity, and Social History," *Games and Economic Behavior,* 10, 122-142.

Camerer, Colin F. (1997) "Progress in Behavioral Game Theory," *J. of Economic Perspectives,* 11, 167-188.

Elster, Jon (1998) "Emotions and Economic Theory," *J. of Economic Literature,* 36; 47-74.

Drèze, Jacques, H. (1987) *Essays on Economic Decisions Under Uncertainty.* Cambridge: Cambridge University Press.

Fishburn, Peter C. (1973) "A Mixture-Set Axiomatization of Conditional Subjective Expected Utility," *Econometrica* 4l: l-25.

Ghirardato, Paolo (2002) "Revisiting Savage in a Conditional World," *Economic Theory,* 20: 83-92.

Grant, Simon, and Edi Karni, (2003) "A Theory of Quantifiable Beliefs," *Journal of Mathematical Econnomic (* forthcoming)

Holmstrom, Bengt R. (1979) "Moral Hazard and Observability," *Bell Journal of Economics*, 10: 74-92.

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

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

Karni, Edi (2003) "Foundations of Bayesian Theory," Working paper, Johns Hopkins University.

Karni, Edi, and David Schmeidler (1980) "An Expected Utility Theory for State-Dependent Preferences." Working Paper 48-80, Foerder Institute for Economic Research, Tel-Aviv University.

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

Karni, Edi, and Philippe Mongin (2000) "On The Determination of Subjective Probability by Choice," *Management Science*, 46: 233 - 248.

Loewenstein, George (2000) "Emotions in Economic Theory and Economic Behavior," *The American Economic Review; Papers and Proceedings. 426-432.*

Luce, Duncan R. and David H. Krantz (1971) "Conditional Expected Utility," *Econometrica,* 39*: 253-271.*

Romer, Paul M. (2000) "Thinking and Feeling," *The American Economic Review; Papers and Proceedings. 439-443.*

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

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

Savage, Leonard, J. (1971) "Elicitation of Personal Probabilities and Expectations," *Journal of the American Statistical Association* 66, 783-801.

Shavell, Stephen (1979) "Risk Sharing and Incentives in Principal and Agent Relationship," *Bell Journal of Economics*, 10: 55-73.

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