Agency Theory with Non-Additive Uncertainty

Paolo Ghirardato[1]
Division of the Humanities
and Social Sciences
California Institute of Technology
Pasadena, CA 91125
E-mail: paolo@hss.caltech.edu

(December 1992; This version:June 1994)

[1] This is a revised version of chapter 3 of my doctoral dissertation at UC Berkeley. I am grateful to Bob Anderson, Marco Li Calzi, Alessandro Lizzeri and audiences at Università Bocconi, at the *Third Colloquia on Economic Research* at IGIER, Milano, and at FUR VII, Oslo, for helpful comments and discussion. The usual disclaimer applies. Financial support from Università Bocconi is gratefully acknowledged.

Abstract

We study the effects on a simple agency problem of assuming that parties display beliefs which are not necessarily represented by additive measures, as will be the case if they are uncertainty averse or if there are unforeseen contingencies. We present the players' problems, prove existence of solutions, and discuss analogies and differences with the standard case in the characteristics of optimal incentive schemes. It is shown that *quality of information*, which cannot be captured in the additive case, can be extremely important for both parties' choices. In fact we discuss improvements in the quality of information, and prove that they are going to be beneficial to the principal in a number of cases. This is not in general true for (the natural generalization of) changes in Blackwell informativeness.

1 Introduction

In a typical agency problem one party, called *principal*, has to hire another, called *agent*, to perform some task. The principal cannot observe the agent's action, but she observes the realization of a random variable (for example her revenues) whose distribution is affected by it. Thus she has to devise a payment structure in order to induce the agent to take the course of action which is best possible for her given the informational constraints. Clearly what the principal is facing is just a statistical inference problem, and it is obvious that the structure of the uncertainty that different actions induce plays a key role in determining how well the principal does.

The agency problem has been at the center of a large and growing literature in the last decade, ever since Holmström's [12] and Grossman and Hart's [10] seminal contributions. All these works employ the standard subjective expected utility (SEU for short) model, and in particular they assume that the beliefs of all parties can be represented by additive measures. While this is the obvious starting point, empirical evidence and casual introspection suggest that this assumption might imply imposing a severe restriction on the generality of the results obtained.

In a classical paper [5], Daniel Ellsberg presented some experimental evidence that in certain situations decision makers make choices that are incompatible with any decision rule that requires the existence of an (additive) probability measure on the set of the possible states of the world, such as SEU. He faced the subjects with one urn containing 90 balls: 30 balls are red while the other 60 are black and yellow in an unknown proportion. Then he asked to rank the following four bets:

1. $ 1000 if a red ball is extracted from the urn, 0 otherwise:
2. $ 1000 if a black ball is extracted, 0 otherwise;
3. $ 1000 if a red or a yellow ball is extracted, 0 otherwise;
4. $ 1000 if a black or a yellow ball is extracted, 0 otherwise.

The majority of the subjects stated the following preferences:

$$1 \succ 2 \qquad 4 \succ 3$$

but this would be impossible if their choices reflected an underlying probability function over the set of states of the world. To see this let $P(r), P(b)$ and $P(y)$ be the probabilities of the event that the ball extracted is, respectively, red, black or yellow. Then $1 \succ 2$ implies $P(r) > P(b)$, while $4 \succ 3$ implies $P(r \cup y) < P(b \cup y)$, and both inequalities cannot hold if P is additive. As Ellsberg observed, the reason for such pattern of choices is obvious: People tend to prefer situations in which the structure of uncertainty is better defined, or, to use Frank Knight's famous distinction [14], there is more *risk* and less *uncertainty*. This is the case for bets 1 and 4, for in those cases the decision maker (henceforth DM) *knows exactly* how many cases are favourable and how many are not. It is also interesting to note that this 'mistake' was done by experts as well (even Savage himself, Ellsberg reports), and that even when they realized the problem many just refused to change their preferences. In general it seems that most DMs prefer to bet on events about which they have better information (either because it is broader or because it is more reliable), and thus display what has been termed an *aversion to uncertainty.*[1]

In recent years decision theorists have offered a possible formal solution to the puzzle of uncertainty aversion. Starting from the seminal work of David Schmeidler ([17], first circulated in 1982) they studied axiomatic models of non-additive representations of uncertainty. For example, these models allow the DM facing Ellsberg's problem to be willing to buy bet 2 for \$ 200 (thus revealing $P(b) = 1/5$, if we assume that her preferences are linear) and bet 3 for \$ 400 (thus revealing $P(r \cup y) = P(b^c) = 2/5$). While this might seem *prima facie* unintuitive, it should be remembered that in Savage's operationalistic approach probabilities are (observable) bet prices, and it seems perfectly reasonable that the amount and quality of information are taken into account when forming those prices.

A more basic criticism of the SEU theory (and of most decision theory, as a matter of fact) is that DMs are never really facing a *given* decision problem. The problem has to be structured, and that is often the hardest part of the exercise of practically applying any decision rule. In particular the DM has to construct the set of states of the world, which in the theory are taken to be "exhaustive descriptions of all aspects relevant to the decision problem". It seems clear that no sensible DM would have any hopes of accomplishing such an immense feat but for some scarcely realistic laboratory examples. In most circumstances she will be aware that the state space she perceives is just a very coarse version of the "true" state space for the problem. Using (in a wide sense) a terminology which has become very fashionable recently, she realizes that there are relevant *unforeseen contingencies*. In such circumstances she might understand that also her perception of actions is incomplete, and it seems reasonable to provide for the possibility that she perceive them as maps from states to *sets* of consequences rather than to singletons. This is discussed and justified in detail in [6] and in works by other authors (e.g., Mukerji [16] and, with a different interpretation, Jaffray and Wakker [13] and Hendon *et al.* [11]). Interestingly it is shown that also in these situations, if preferences satisfy some axioms then they can be represented via non-additive beliefs.

The discussion above provides the motivation for this paper. Our objective is to apply the decision rule which arises in both the families of models sketched above, to a very simple agency problem (with finitely many revenue levels for the principal), and to see the effects of this generalization[2] on the solutions to the agency problem (as compared, say, to the results presented in Grossman and Hart [10]). More importantly, this generalization allows us to analyze a totally new side of the information problem: How important is the ambiguity of the stochastic structure of returns? Is the principal going to benefit from a reduction in the

[1]It should be stressed at this point that this has no relation to *risk* aversion. A DM can be risk averse without being uncertainty averse, and *vice versa*.

[2]As it can be easily seen, the model we study has SEU as a special case.

non-additivity of beliefs?[3]

As it turns out, the introduction of non-additivity has no disruptive consequences on standard results. We shall see that, while with non-additive priors it is no more possible to separate the principal's problem in two stages, it is still possible to prove that such a problem has a solution. However differences in the quality of information about the agent's actions *do* play an important role in the principal's choice, as a result and some examples will testify.

The second question raises the more general issue of the comparative statics of information. As it seems intuitively clear, improving the quality of information or the specification of the state space reduces the cost of implementing some actions and so it often increases principal's profits. On the other hand the fact that information is symmetrically distributed will imply that this will not happen always, as it will depend on *which* actions parties become more informed about. There are, of course, additive notions of informativeness which are complementary to the one just discussed and are well-known to economists, for instance Blackwell informativeness. We will see that some well-known comparative statics results using this notion do not easily extend to non-additive beliefs, at least with its most natural generalizations to this larger class of measures.

The paper is organized as follows. In part 2 we briefly sketch formally the two decision models outlined above and present and explain the decision rule employed in the rest of the paper. Section 3.1 introduces the agency model and our assumptions. The principal's problem is outlined and discussed in section 3.2, and the existence of a solution to it is proved in section 3.3. We analyze more in depth the structure and features of such solution in section 3.4. In part 4 we discuss the comparative statics of information. Section 4.1 analyzes the effects of changes in the non-additivity of beliefs, after discussing how such a change should be properly defined. Likewise section 4.2 starts with a discussion about the proper generalization of the concept of Blackwell informativeness and then proceeds to analyze the comparative statics.

2 Models of Non-Additive Uncertainty

Let Ω be a set of states of the world and Σ be an algebra of subsets of Ω. A set-function $\nu : \Sigma \to \mathbf{R}$ is called a (normalized) *capacity* if it satisfies:

(i) $\nu(\emptyset) = 0,\ \nu(\Omega) = 1$,

(ii) $\forall A, B \in \Sigma : A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$.

Note that (i) and (ii) imply that the range of ν is contained in $[0, 1]$. A capacity is called *convex* if in addition to (i)-(ii) it satisfies

(iii) $\forall A, B \in \Sigma :\ \nu(A \cup B) + \nu(A \cap B) \geq \nu(A) + \nu(B)$.

[3]It is important to point out from the outset that, following the standard literature on agency theory, we will throughout assume that there is a common prior, so that the non-additive beliefs will be shared by both principal and agent, even though we believe that some more insights might come from relaxing this assumption.

It is called a *probability* if (iii) holds with equality for every A and B. Obviously every probability is a convex capacity, but not *vice versa*.

To obtain an integral representation of preferences which parallels the standard SEU representation we need a concept of integral with respect to a capacity. Schmeidler [17] suggested to use the integral proposed by Gustave Choquet in [3]. If $a : \Omega \to \mathbf{R}$ is a bounded Σ-measurable function and ν is a capacity (not necessarily convex) on Ω we define the *Choquet integral* of a with respect to ν to be the number

$$\int_\Omega a(\omega)\, d\nu(\omega) = \int_0^\infty \nu(\{\omega \in \Omega : a(\omega) \geq \alpha\})\, d\alpha + \int_{-\infty}^0 [\nu(\{\omega \in \Omega : a(\omega) \geq \alpha\}) - 1]\, d\alpha$$

where the integrals are taken in the sense of Riemann. In particular, if Ω is finite (that is, $\Omega = \{\omega_1, \ldots, \omega_n\}$) and $a(\omega_1) \geq a(\omega_2) \geq \ldots \geq a(\omega_n)$, then

$$\int_\Omega a(\omega)\, d\nu(\omega) = \sum_{i=1}^{n-1} (a(\omega_i) - a(\omega_{i+1}))\nu(\{\omega_1, \ldots, \omega_i\}) + a(\omega_n).$$

Notice that since the integrands are monotone, the Choquet integral always exists, and if ν is a probability it reduces to a standard Lebesgue integral. Also observe that if $a(\omega_i) = a(\omega_{i+1})$, then if we relabel the states of the world, so that the labels assigned to state i and $i+1$ are swapped and all the others remain the same, the value of the integral will remain unchanged. Henceforth all the integrals will be taken in the sense of Choquet, except where otherwise noted.

When integrating on a *convex* capacity, the Choquet integral behaves in a very "pessimistic" manner. To see what this means let us recall a very nice equivalence result due to Schmeidler [17, Proposition]. Given a convex capacity ν define the *core* of ν to be the set

$$\mathcal{C}(\nu) = \{P : \Sigma \to \mathbf{R} : P \text{ is additive}, P(\Omega) = \nu(\Omega), \forall E \in \mathcal{S}, P(E) \geq \nu(E)\},$$

i.e., the set of probabilities on Σ which *dominate* ν. It is a well-known result from cooperative game theory that the core of a convex capacity is non-empty. If $a : \Omega \to \mathbf{R}$ is bounded and Σ-measurable then

$$\int_\Omega a(\omega)\, d\nu(\omega) = \min_{P \in \mathcal{C}(\nu)} \int_\Omega a(\omega)\, dP(\omega). \tag{1}$$

Thus we can interpret a convex capacity as the lower envelope of a *set* of probabilities, and the Choquet integral as an operator which integrates (in the standard sense) a function a using the "worst possible" probability, the one which minimizes the expectation of a. In other words a DM whose preferences are representable by a Choquet integral with respect to a convex capacity behaves as if she had a set of priors $\mathcal{C}(\nu)$ and she maximized the minimum expected utility on $\mathcal{C}(\nu)$.

Schmeidler presented in [17] a set of axioms which yield as representation a Choquet integral with respect to a capacity. In his model the objects of choice, called *acts*, are functions from the state space Ω into the set $\mathcal{P}$ of all *lotteries* (i.e., probabilities with a finite support)[4] on some set of consequences $\mathcal{X}$. The set of all acts is denoted $\mathcal{F}$. The DM's preferences $\succeq$ on $\mathcal{F}$ satisfy Schmeidler's axioms if and only if there is a utility function $U : \mathcal{P} \to \mathbf{R}$, unique up to a positive affine transformation and affine[5] on $\mathcal{P}$, and a unique capacity ν on Σ such that for every f, $g \in \mathcal{F}$,

$$f \succeq g \Longleftrightarrow \int_\Omega U(f(\omega))\, dv(\omega) \geq \int_\Omega U(g(\omega))\, dv(\omega). \tag{2}$$

If, moreover, the DM's preferences satisfy an additional axiom called "uncertainty aversion", then ν will be convex. Intuitively, an uncertainty averse DM prefers substituting objective (risky) mixing for subjective (uncertain) mixing, hence the name given to this property. For every act f we can define the distribution on $\mathcal{X}$ as follows: For $X \subseteq \mathcal{X}$, let

$$\nu_f(X) = \nu(\{\omega \in \Omega : f(\omega) \in X\}) = \nu(f^{-1}(X)). \tag{3}$$

Notice that if ν is a convex capacity then ν_f will be a convex capacity for all f. Obviously we have

$$\int_\Omega U(f(\omega))\, d\nu(\omega) = \int_\mathcal{X} U(x)\, d\nu_f(x),$$

so we can equivalently represent the DM's preferences with the integral on the distribution induced by each act.

As we saw earlier, another situation in which non-additive beliefs might be generated is when the DM's perception of the state space Ω is "coarse", and she is aware of some incompleteness in her perception. This problem has been analyzed theoretically in [6]. Suppose that the DM only perceives as state space a *partition* Π of Ω. If $\mathcal{X}$ is the finite set of possible outcomes, then she is assumed to perceive acts in $\mathcal{F}$ according to what we call a *perception correpondence* $\mathbf{C} : \mathcal{F} \times \Pi \to 2^{\mathcal{X}}$, that is, each act $f \in \mathcal{F}$ is a *correspondence* from Π to $\mathcal{X}$ (to simplify notation we write $f_c(\pi)$ rather than $\mathbf{C}(f, \pi)$). As we anticipated in the introduction, the interpretation of this is that, even though she does not understand why, the DM might realize that act f can yield more than one possible outcome in some (perceived) state π.[6]

If the DM's preferences $\succeq$ on $\mathcal{F}$ satisfy some natural generalizations of Savage's axioms to this set-up, then her beliefs on Π can be represented by a (finitely additive) probability measure P defined on all subsets of Π. For any act $f \in \mathcal{F}$ let ν_f be the "distribution" on $\mathcal{X}$ defined as follows, for every $X \subseteq \mathcal{X}$,

$$\nu_f(X) = P(\{\pi \in \Pi : f_c(\pi) \subseteq X\}). \tag{4}$$

[4] In other words there is an independent randomizing device with "objectively known" probabilities. Schmeidler's result has been later generalized to a framework without objective lotteries (à la Savage) by Gilboa [7] and Wakker [19].

[5] That is, if $p, q \in \mathcal{P}$ and $\alpha p \oplus (1-\alpha)q$ is their mixture with weight $\alpha \in [0,1]$ then $U(\alpha p \oplus (1-\alpha)q) = \alpha U(p) + (1-\alpha)U(q)$. In other words, the DM behaves as an expected utility maximizer with respect to lotteries.

[6] Say because she has taken act f more than once in similar problems and she observed various different outcomes in state π.

Notice that if f_c is a function then ν_f is a probability distribution, and in general it is a convex capacity on $2^{\mathcal{X}}$. Actually more is true: ν_f satisfies a property stronger than (iii).

(iii') For every $n > 0$ and every collection $X_1, \ldots, X_n \in 2^{\mathcal{X}}$

$$\nu_f(\cup_{i=1}^{n} X^i) \geq \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ I \neq \emptyset}} (-1)^{|I|+1} \nu_f(\cap_{i \in I} X_i)$$

where $|I|$ is the cardinality of set I.

That is, ν_f is what (after Shafer [18]) is called a *belief function*. In [6] it is proved that the afore-mentioned axioms plus another axiom (which intuitively depicts a strongly pessimistic attitude) imply that there is a utility function $U : \mathcal{X} \to \mathbf{R}$, unique up to a positive affine transformation, such that for every $f,\, g \in \mathcal{F}$,

$$f_c \succeq g_c \Longleftrightarrow \int_{\mathcal{X}} U(x)\, d\nu_f(x) \geq \int_{\mathcal{X}} U(x)\, d\nu_g(x). \tag{5}$$

Thus the DM's preferences can again be represented by a Choquet integral with respect to the distributions on $\mathcal{X}$.

In a sense these two models describe two complementary sources of the non-additivity of the distributions ν_f. There is on one hand the vagueness of the information on which probability judgements on the (perceived) state space are based. And then there is the (perceived) indeterminacy of the results of acts, due to incomplete specification of the state space.

3 The Agency Problem: Optimal Incentive Schemes and their Characteristics

3.1 The Model

The set-up of our model is fairly standard. There are two individuals, one, the principal, needs (for some unmodelled reasons) to hire the other, the agent, to perform a specific task. We let Ω be a set of states of the world, $\mathcal{X}$ a finite set of *outcomes*[7] (in which the principal is interested) and $\mathcal{F}$ be a finite set of *actions* that the agent can take. Specifically, an action is a measurable function $f : \Omega \to \mathcal{X}$, i.e., for every $\omega \in \Omega$ $f(\omega)$ is an outcome for the principal. We let n be the number of elements of $\mathcal{X}$, that is, $\mathcal{X} = \{x_1, \ldots, x_n\}$.

It is assumed that the principal cannot observe the action f that the agent truly undertakes, therefore she can only make her payment (in amounts of money) to the agent depend on the outcome $x \in \mathcal{X}$ that she obtains from the agent's activity. Such a payment we call an *incentive scheme*, formally it is a function $Y : \mathcal{X} \to \mathbf{R}$. Given the stochastic structure

[7] We could equivalently assume that the set of states of the world is finite, or that all actions are *simple*, i.e., they have finite range. The reason for this particular choice will become apparent shortly.

of the problem, in general the principal will not be able to deduce the agent's action by the outcome (but see below).

We will assume that both players' preferences can be represented by a Choquet integral with respect to the distribution function induced on $\mathcal{X}$ by each action, and that this is *common* for both. As we saw in section 2 this could be due to either uncertainty aversion or to incomplete knowledge of the state space. In the former case players have a common convex capacity ν on the state space Ω, which induces the non-additive distribution ν_f on $\mathcal{X}$ for every action f. In the latter case both players perceive the same partition Π of Ω, they have the same perception correspondence $\mathbf{C}$ (that is, they agree on the way actions are constructed) and they have a common (additive) prior P on Π. Thus they agree that action $f \in \mathcal{F}$ induces the distribution ν_f defined in equation (4). Following Grossman and Hart ([10], assumption A3), we will assume that $\nu_f(x_i) > 0$ for each $f \in \mathcal{F}$ and $i = 1, \ldots, n$, that is, every action induces a positive probability of every outcome (this is what in the literature is called the "no shifting support condition").

The agent's utility function $V : \mathbf{R} \times \mathcal{F} \to \mathbf{R}$ has the usual separable form $V(y, f) = u(y) - c(f)$, where $u : \mathbf{R} \to \mathbf{R}$ is strictly increasing, concave, and represents the utility from money income (thus the agent can be risk-averse). $c : \mathcal{F} \to \mathbf{R}$ is non-negative and it can be interpreted as the *disutility* of taking action f. Therefore, given that the principal is employing the incentive scheme Y, the agent chooses $f \in \mathcal{F}$ so as to maximize

$$V_Y(f) \equiv \int_{\mathcal{X}} u(Y(x))\, d\nu_f(x) - c(f). \tag{6}$$

The agent has reservation utility $\underline{u}$, i.e., he will accept the job he is offered only if his expected utility is higher than $\underline{u}$. Without loss of generality we assume that $\underline{u} \geq 0$ and that there is an action $l \in \mathcal{F}$ such that $c(l) < c(f)$ for each $f \neq l$, $f \in \mathcal{F}$. l is called the *least cost action.*

The principal is risk-neutral, so that she chooses $Y \in \mathbf{R}^n$ and $f \in \mathcal{F}$ in order to maximize

$$\int_{\mathcal{X}} (x - Y(x))\, d\nu_f(x)$$

subject to the constraint that Y implements f (to be explained below).

3.2 The Principal's Problem

Suppose that principal wants to induce agent to take action $f \in \mathcal{F}$ by paying him the incentive scheme Y. Then she is going to reach her objective only if Y satisfies the following definition.

Definition 1 *Incentive scheme Y implements f if the following constraints are satisfied*

(IC) $\int_{\mathcal{X}} u(Y(x))\, d\nu_f(x) - c(f) \geq \int_{\mathcal{X}} u(Y(x)) d\nu_g(x) - c(g), \qquad \forall g \in \mathcal{F};$

(IR) $\int_{\mathcal{X}} u(Y(x))\, d\nu_f(x) - c(f) \geq \underline{u}.$

Define $Y_f = \{Y \in \mathbf{R}^n : Y \text{ implements } f\}$. Obviously l can always be implemented efficiently with the constant incentive scheme $Y \equiv \bar{y}$, where[8]

$$\bar{y} = u^{-1}(\underline{u} + c(l)) \equiv C^{FB}(l),$$

as then both (IR) and all the (IC) are obviously satisfied. For $Y \in Y_f$, let

$$\phi_f(Y) = \int_{\mathcal{X}} (x - Y(x)) d\nu_f(x)$$

so if we define $\Psi(f) \equiv \max_{Y \in Y_f} \phi_f(Y)$ we have that the principal's problem is to

$$\max_{f \in \mathcal{F}} \Psi(f). \tag{7}$$

Following Grossman and Hart, we shall call *second best action* an f which solves (7) and *second best incentive scheme* a Y which implements such action. In general, we shall define *optimal* an incentive scheme which maximizes $\phi_f(\cdot)$ for some f.

Example 1 To illustrate the difference with the additive case, let us consider a problem in which the principal distinguishes only two outcomes, a "success" and a "failure", i.e., $\mathcal{X} = \{x_1, x_2\}$, where $x_1 > x_2$. For every f and every $i = 1, 2$ let $\nu_{fi} = \nu_f(x_i)$. Given the particular structure of the Choquet integral, the exact formula for (IC) and (IR) will then depend on whether $z_1 \geq z_2$, where $z_i = u(Y(x_i))$ (or, equivalently since u is strictly increasing, $y_1 \geq y_2$, where $y_i = Y(x_i)$) or $z_1 < z_2$ ($y_1 < y_2$).

If $z_1 \geq z_2$ then we can write

$$\text{(IC-1)} \qquad \nu_{f1} z_1 + (1 - \nu_{f1}) z_2 - c(f) \geq \nu_{g1} z_1 + (1 - \nu_{g1}) z_2 - c(g), \qquad \forall g \in \mathcal{F};$$

$$\text{(IR-1)} \qquad \nu_{f1} z_1 + (1 - \nu_{f1}) z_2 - c(f) \geq \underline{u}.$$

If $z_1 < z_2$ then we can write

$$\text{(IC-2)} \qquad (1 - \nu_{f2}) z_1 + \nu_{f2} z_2 - c(f) \geq (1 - \nu_{g2}) z_1 + \nu_{g2} z_2 - c(g), \qquad \forall g \in \mathcal{F};$$

$$\text{(IR-2)} \qquad (1 - \nu_{f2}) z_1 + \nu_{f2} z_2 - c(f) \geq \underline{u}.$$

Notice that the equations coincide if ν_f and ν_g are additive, but they are different if either is strictly non-additive. Thus the agent will always discount the probability of the event he considers to be best, as we mentioned earlier.

For future reference, it will be useful to partition the set Y_f in three subsets as follows

$$Y_f = Y_f^{<} \cup Y_f^{>} \cup Y_f^{>>}$$

[8] For every f, $C^{FB}(f)$ stands for first best cost of implementing f, i.e., the cost the principal would sustain in the (first best) situation in which she could observe the agent's action.

where

$$\begin{aligned} Y_f^< &\equiv \{Y \in Y_f : y_1 < y_2\} \\ Y_f^> &\equiv \{Y \in Y_f : y_2 + x_1 - x_2 \geq y_1 \geq y_2\} \\ Y_f^{>>} &\equiv \{Y \in Y_f : y_1 > y_2 + x_1 - x_2\} \end{aligned}$$

The set $Y_f^<$ is the set of non-monotonic incentive schemes, in the sense that payments to the agent do not respect the order of the revenues to the principal. The set $Y_f^{>>}$ is the set of monotonic incentive schemes in which the principal pays so much when the good outcome occurs that she is going to have lower *profits* than if the bad outcome occurs.

Now, for $f \in \mathcal{F}$ let

$$\begin{aligned} R_f &= \nu_{f1} x_1 + (1 - \nu_{f1}) x_2, \\ R_f^{>>} &= (1 - \nu_{f2}) x_1 + \nu_{f2} x_2. \end{aligned}$$

Analogously, for every $y \in \mathbf{R}^2$ let

$$\begin{aligned} C_f(y) &= \nu_{f1} y_1 + (1 - \nu_{f1}) y_2, \\ C_f^{>>}(y) &= (1 - \nu_{f2}) y_1 + \nu_{f2} y_2. \end{aligned}$$

Finally, define $C_f \equiv \min_{Y \in (Y_f \setminus Y_f^{>>})} C_f(Y)$ and $C_f^{>>} \equiv \min_{Y \in (Y_f^{>>})} C_f^{>>}(Y)$. Obviously $R_f = R_f^{>>}$, and $C_f = C_f^{>>}$ when ν_f is additive.

We can now see that the principal's problem is to maximize

$$\Psi(f) = \max[(R_f - C_f),\ (R_f^{>>} - C_f^{>>})].$$

In particular it is not hard to see that $\Psi(l) = R_l - u^{-1}(\underline{u} + c(l))$. Note that differently from the case in which ν is additive, it is not possible to formulate the principal's problem of implementing f simply as a cost-minimization problem: The principal's revenue function depends on the incentive scheme she employs.

This phenomenon is by no means specific to the two-outcome case, and it is a result of the non-additivity of the measure ν_f, which prevents the usual separation of the principal's objective function $\phi_f(\cdot)$ in revenues (which do not depend on the incentive scheme used) and costs. Since the principal's probability assessments will change with the incentive schemes she offers, her expected revenues will change as well. By a well-known fact on Choquet integrals (see Schmeidler [17, Proposition])

$$\phi_f(Y) = \int_{\mathcal{X}} (x - Y(x))\, d\nu_f(x) \geq \int_{\mathcal{X}} x\, d\nu_f(x) + \int_{\mathcal{X}} -Y(x)\, d\nu_f(x)$$

and strict inequality might hold for non-additive ν_f.

This notwithstanding, many results of the standard additive analysis are still valid in the more general case, the following being an example.

Remark 1 In the additive case it is well-known that if the agent is risk-neutral, then the principal will be able to obtain her first-best profit by choosing the incentive scheme Y defined by $y_i = x_i - P$ where

$$P = \int_{\mathcal{X}} x \, d\nu_{f^*}(x) - C^{FB}(f^*)$$

and $f^* = \arg\max_{f \in \mathcal{F}} \{\int_{\mathcal{X}} x \, d\nu_f(x) - C^{FB}(f)\}$. One easily sees that the result can be generalized to allow for non-additive beliefs, as long as they are shared by both players. What is important in this case is that the principal can induce the agent to maximize her utility since, thanks to the choice of Y, they have the same objective function. Now, it can be seen that since $(y_1, \ldots, y_n)$ and $(x_1, \ldots, x_n)$ differ only by a constant the non-additivity of beliefs does not matter. However it should be kept in mind that counterexamples to this result can be easily constructed if we allow heterogenous priors (even if they are additive).

3.3 Existence of Solutions to the Principal's Problem

The non-separability mentioned above might induce the reader to wonder whether a solution to the principal's problem always exists under the conditions stated so far. First of all we have seen earlier that, exactly as in the standard case, the least cost action can always be implemented at first best cost with expected profit $\Psi(l) = R_l - C^{FB}(l)$. Now, if $Y_f = \emptyset$ then $\Psi(f) = -\infty$ so to prove existence we only have to show that for each $f \in \mathcal{F}$, $Y_f \neq \emptyset$ implies that $\phi_f(Y)$ attains a maximum in Y_f. Given that, the fact that $\mathcal{F}$ is finite is enough to imply that the principal's problem has a solution. We begin by proving a well-known result.

Lemma 1 *For $f \in \mathcal{F}$, suppose that $Y \in Y_f$ is such that (IR) holds with inequality, i.e.*

$$\int_{\mathcal{X}} u(Y(x)) \, d\nu_f(x) - c(f) > \underline{u}$$

then $\exists \bar{Y} \in Y_f$ such that $\phi_f(\bar{Y}) > \phi_f(Y)$.

Proof: Let $\epsilon > 0$ be such that

$$\int_{\mathcal{X}} [u(Y(x)) - \epsilon] \, d\nu_f(x) - c(f) > \underline{u}$$

Given the hypothesis on u it is possible to find $\bar{Y} = (\bar{y}_1, \ldots, \bar{y}_n)$ such that $u(\bar{y}_i) = z_i - \epsilon$, $i = 1, 2, \ldots, n$. Also, monotonicity of u implies that for each i, $\bar{y}_i > y_i$, so that $\phi_f(\bar{Y}) > \phi_f(Y)$. □

Theorem 2 *For every $f \in \mathcal{F}$, $\phi_f(Y)$ attains a maximum in Y_f.*

Proof: $\phi_f : \mathbf{R}^n \to \mathbf{R}$ is continuous over Y_f, hence it is only necessary to prove that in her search for optimal incentive schemes, the principal can restrict her attention to a compact subset of Y_f, for then the result follows from Weierstrass's theorem. By lemma 1 only incentive schemes that satisfy (IR) with equality can be optimal, hence define

$$\mathcal{Y}_f \equiv \{Y \in Y_f : \int_{\mathcal{X}} u(Y(x))\, d\nu_f(x) - c(f) = \underline{u}\}$$

If u is linear then the result in remark 1 applies, so that the statement is certainly true. So assume that u is nonlinear.

Let Π the set of all the $n!$ permutations of $\{1, 2, \ldots, n\}$, i.e. the set of all the one-to-one functions $\pi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. One can obviously partition $\mathbf{R}^n$ in $n!$ sets

$$R^\pi \equiv \{y \in \mathbf{R}^n : \forall i, j = 1, 2, \ldots, n,\ i \neq j,\ \pi(i) < \pi(j) \Rightarrow y_i \geq y_j\}$$

Let $\mathcal{Y}_f^\pi = \mathcal{Y}_f \cap R^\pi$ and let $\mathcal{Z}_f^\pi$ be its image through u. It is sufficient to prove that, for every permutation π, there is a compact subset of $\mathcal{Y}_f^\pi$ such that only the schemes belonging to it are optimal, for it is clear that, once that is proved, the fact that there only finitely many such permutations yields the result.

Without any loss of generality we can renumber the outcomes so that $\pi = id$, where id is the identity function[9]. Let $p_i = x_i - y_i$ be the profit of the principal in state i. Depending on which incentive scheme Y she is paying, principal's profit will ordered according to some permutation π_Y so that we can write $p_{1,\pi_Y} \geq p_{2,\pi_Y} \geq \cdots \geq p_{n,\pi_Y}$.

If $\mathcal{Y}_f = \emptyset$ then the statement follows, so assume that $\mathcal{Y}_f \neq \emptyset$. Let

$$\begin{aligned} m_i &= \nu_f(\{x_1, \ldots, x_i\}) - \nu_f(\{x_1, \ldots, x_{i-1}\}), \\ m_i^{\pi_Y} &= \nu_f(\{x_{1,\pi_Y}, \ldots, x_{\pi_Y(i),\pi_Y}\}) - \nu_f(\{x_{1,\pi_Y}, \ldots, x_{\pi_Y(i)-1,\pi_Y}\}). \end{aligned}$$

As ν_f is monotone, all the numbers above are non-negative. In fact, one can interpret m_i as the *probability* of getting y_i under the permutation $\pi = id$, for one easily sees that:

$$\int_{\mathcal{X}} u(Y(x))\, d\nu_f(x) = \sum_{i=1}^n z_i\, m_i.$$

The interpretation of m_i^π is analogous. Consider the case in which $m_i = m_i^{\pi_Y}$, i.e. the principal and the agent use the same probabilities. Following Grossman and Hart ([10], proposition 1) we can apply a result in Bertsekas [1] to show that if $\{z^k\}_{k=1}^\infty$ is an unbounded sequence in[10] $\mathcal{Z}_f$, i.e., for every $M > 0$ can find k such that $||z_k|| > M$, given that u^{-1} is nonlinear and convex, and that $\nu_f(x_i) > 0$ for every i (which by the convexity of ν_f implies that $m_i > 0$), then

$$\lim_{k\to\infty} \sum_{i=1}^n (x_i - u^{-1}(z_i^k))\, m_i = -\infty.$$

[9] And, to simplify notation, in the rest of the proof we shall omit the id superscript.

[10] Observe that for every k the fact that $z_k \in \mathcal{Z}_f$ implies that $\sum_{i=1}^n z_i^k\, m_i$ is constant.

But by the maxmin property of Choquet integral given in equation (1) , we also have that for every $Y \in \mathbf{R}^n$

$$\int_{\mathcal{X}} (x - Y(x))\, d\nu_f(x) \leq \sum_{i=1}^{n} (x_i - y_i)\, m_i$$

so that we conclude that

$$\limsup_{k\to\infty} \int_{\mathcal{X}} (x - Y^k(x))\, d\nu_f(x) \leq \lim_{k\to\infty} \sum_{i=1}^{n} (x_i - y_i^k)\, m_i = -\infty,$$

where $y_i^k = u^{-1}(z_i^k)$. So we can artificially put a bound on the constraint set $\mathcal{Y}_f$, which is what we wanted to prove. □

Remark 2 We can exactly use the technique employed by Grossman and Hart ([10], proposition 1) to show that the existence result holds also when the action space $\mathcal{F}$ is not finite but it is compact in some topology. Let $\{f_k\}_{k=1}^{\infty}$ be a sequence converging to f. Assume that $\Psi(f_k) \to M$. If $M = -\infty$ then we certainly have $\Psi(f) \geq \lim_{k\to\infty} \Psi(f_k)$. If $M > -\infty$ then by the result of Bertsekas, if Y^k is the optimal incentive scheme for f_k then the sequence $\{Y^k\}_{k=1}^{\infty}$ is bounded (otherwise $\Psi(f_k) \to -\infty$). Let Y be a limit point, then $Y \in Y_f$ and thus

$$\Psi(f) \geq \phi_f(Y) = \lim_{k\to\infty} \Psi(f_k).$$

This proves that Ψ is upper semicontinuous, so that again the result follows from Weierstrass's theorem.

3.4 On the Structure of Optimal Incentive Schemes and Actions

We have already seen some characteristics of optimal incentive schemes. Lemma 1 told us that if the principal is employing an optimal incentive scheme the agent will never receive an expected utility higher than $\underline{u}$. Analogously we could prove (this is Grossman and Hart's proposition 6) that at least one of the incentive compatibility constraints will be binding if f is not the least cost action.

An interesting problem is whether the incentive scheme will be monotonic, in the sense that higher profits for the principal will correspond with higher payments to the agent. Grossman and Hart [10] studied the problem extensively (with additive uncertainty) and they concluded that if there are more than two outcomes it will not in general be the case that the incentive scheme is monotonic. On the other hand their proposition 4 implies that monotonicity is assured in the case with two outcomes. The following example shows that allowing non-additive uncertainty invalidates this result.

Example 1 (continued) Suppose that there are only two actions: f and l, and that ν_f and ν_l are defined as follows

$$\nu_{f1} = 0.3 \qquad \nu_{f2} = 0.6$$

$$\nu_{l1} = 0.27 \qquad \nu_{l2} = 0.3$$

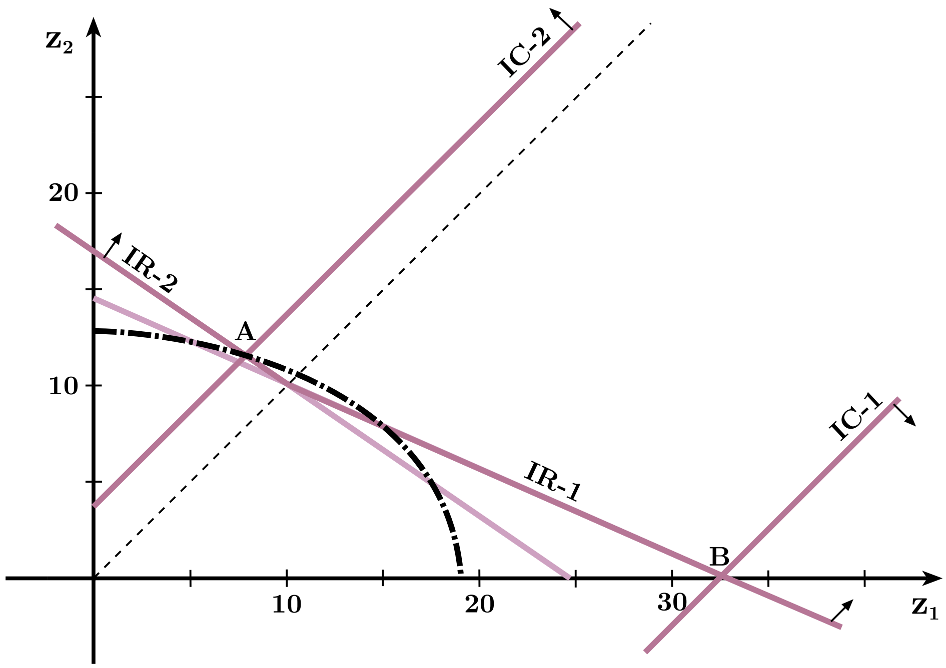

Figure 1: Example 1

Outcomes are $x_1 = 1000,\ x_2 = 0$, while costs are $c(f) = 10,\ c(l) = 9$. Agent's utility function for income is $u(y) = y^{\frac{1}{2}}$ for $y \geq 0$ (the behaviour of u for $y < 0$ is not important). $\underline{u} = 0$, i.e. agent has zero reservation utility.

The principal's problem is depicted in figure 1. Differently from the additive case here we see that the individual rationality constraint IR is only piecewise linear, as it changes slope when it intersects the 45° line (which is represented by a dashed line). Also there are two different incentive compatibility constraints, one on each side of the bisectrix. Thus $Y_f^<$ contains all the points above the IR-2 and to the left of the IC-2 constraints, while $Y_f \setminus Y_f^<$ contains all the points above IR-1 and to the right of IC-1. The dotted curve is part of the principal's indifference curve passing through point A. The latter also changes slope (when entering $Y_f^{>>}$), but as this happens far below the z_1 axis, it is not represented.

As we can see in the figure, f can at best be implemented with the incentive scheme $B \in (Y_f \setminus Y_f^<)$ for a cost[11] of $C_f(B) = 0.3(\frac{100}{3})^2 \approx 333.33$, while there is a $\hat{Y} \in Y_f^<$, namely A, such that $C_f(A) = 0.3(8)^2 + 0.7(11.33)^2 \approx 109.11$. Moreover, one can check that the expected profit of f is higher than that of l, so that f is a second best action.

The reason for this result is that, differently from what can happen in the additive case, action f is *less non-additive* than l, therefore it yields higher expected revenue for the princi-

[11]It is easy to see that the principal would not benefit from offering an incentive scheme $Y \in Y_f^{>>}$. Thus we can restrict our attention to the subset of Y_f where the principal's revenue is constantly equal to R_f.

pal. Also, given that the difference in the probabilities of x_2 is much larger than the difference in the probabilities of x_1, in a sense x_2 is a better signal that the agent has undertaken action f, and thus the principal prefers to pay the agent more in such a case.

In a similar vein, we can prove that if an action f is such that the agent has another action g which does not cost more and induces a distribution on $\mathcal{X}$ which is "similar" but less non-additive,[12] then f cannot be implemented.

Theorem 3 *Suppose that $f \in \mathcal{F}$ is such that $\exists g \in \mathcal{F}$ for which either of the following is true:*

(i) $c(f) > c(g)$ and $\nu_g \geq \nu_f$ (i.e., for every $X \subseteq \mathcal{X}$, $\nu_g(X) \geq \nu_f(X)$);

(ii) $c(f) \geq c(g)$, $\nu_g \geq \nu_f$ and $\nu_g(x) > \nu_f(x)$ for every $x \in \mathcal{X}$;

then $Y_f = \emptyset$, i.e., f is not implementable.

Proof: It is easy to see that if $\nu_g \geq \nu_f$ then $\int_{\mathcal{X}} a(x)\, d\nu_g(x) \geq \int_{\mathcal{X}} a(x)\, d\nu_f(x)$ for every function $a : \mathcal{X} \to \mathbf{R}$. Hence under either condition there can be no Y which satisfies (IC). □

Thus the quality of information players have regarding an action, as reflected in the non-additivity of the distribution it induces on $\mathcal{X}$, plays an important role in determining whether it is implementable or not. It does also affect the action's appetibility for the principal, as the following example shows.

Example 2 This is a variation on an example by David Kreps, presented in his textbook on microeconomics [15, pp.601-603]. There are two outcomes $x_1 = \$10$ and $x_2 = \$0$, the utility function of the agent is $u(y) = ln(y)$ and $\underline{u} = 0$. There are three actions, f, g and h, their respective costs of effort to the agent are $c(f) = 0$, $c(g) = 0.1$ and $c(h) = 2.27$, and they induce the following distributions:

$$\begin{aligned} \nu_{f1} &= 0.1 & \nu_{f2} &= 0.9 \\ \nu_{g1} &= 0.15 & \nu_{g2} &= 0.85 \\ \nu_{h1} &= 0.94 & \nu_{h2} &= 0.01. \end{aligned}$$

Notice that f and g are additive. The only change we made to Kreps's example is that we made h slightly non-additive: his h gave a 0.99 chance of obtaining x_1. He shows that when only f and g are available the principal will choose to implement f, the least cost action, with the flat incentive scheme of \$1, for an expected profit of 0, since the difference in the distributions of f and g is not large enough to offset the difference in cost. On the other hand, when *his* h is also available the principal will choose to implement it, even though its cost to the agent is very high. Intuitively the reason is that h gives a statistical pattern

[12] Either because it is felt to be less dependent on unspecified contingencies or because the agent has a better quality of information. In section 4.1, when discussing the comparative statics of information, we explain this in greater detail.

of outcomes which is much different from the other actions, so that it is less costly for the principal to distinguish it.

But when h gives the distribution above things change. One can check that in such a case the optimal incentive scheme for implementing h is $(13.32, 0.71)$, which gives the principal an expected profit of -3.29. Hence in this case the principal will still prefer to implement f. The non-additivity of the induced distribution, albeit slight, is enough to counterbalance the palatability of h due to its different stochastic structure.

4 The Comparative Statics of Information

A natural question in agency problems is how are the choice of second best action and the principal's profits[13] going to be affected by changes in the information on the stochastic relation between actions and outcomes. The main novelty of our framework is that in addition to standard additive notions of changes in informativeness, we can discuss changes in the players' confidence on their state of information, as formally reflected by changes in the non-additivity of their beliefs.

4.1 Changes in Non-Additivity

We want to analyze what happens when the extent of the non-additivity of the distributions ν_f decreases in some way.[14] We identify two main justifications for this: 1) The (confidence about the) quality of the information on some actions increases, reducing uncertainty aversion; 2) The perception of the state space Π becomes more refined, in the sense that the old Π is seen as a partition of the new Π. How are these changes going to be represented formally?

In case 1 what we are looking for is a global measure of the uncertainty aversion displayed by a convex capacity. Local measures of uncertainty aversion have been offered by Dow and Werlang (DW, [4]) and Gilboa and Schmeidler [8]. Given a capacity ν defined on an algebra Σ of subsets of some state space Ω, for every $T \in \Sigma$ let $c_\nu(T) = 1 - \nu(T) - \nu(\Omega \setminus T)$. It can be seen that if ν is convex, c_ν will be non-negative, and it will be identically zero if and only if ν is additive. So DW proposed the following intuitive condition: We say that μ displays higher uncertainty aversion than ν at $T \subseteq \Omega$, $c_\nu(T) \leq c_\mu(T)$, or equivalently

$$\nu(T) + \nu(\Omega \setminus T) \geq \mu(T) + \mu(\Omega \setminus T). \tag{8}$$

This is made global as follows: μ displays (strictly) higher uncertainty aversion than ν if (8) holds for every $T \in \Sigma$ with strict inequality holding for at least one T.

While it is clearly a very natural quantification of the amount of non-additivity of a capacity, this definition suffers, in our viewpoint, of a serious problem. Decreases in uncertainty aversion in this sense can be associated with drastic changes of the likelihood ratios of events.

[13] By lemma 1, whatever beliefs are, the agent is always going to receive utility $\underline{u}$ in an optimal incentive scheme.

[14] Obviously, we can analyze *increases* in non-additivity just by reversing our results.

That is, it is possible that the players decide that their previous assessments of likelihood were mistaken and, while increasing the total weight allocated between T and its complement $\Omega \setminus T$ (as equation (8) requires), alter their beliefs significantly. Using a (weak) analogy with consumer theory we could say that a change of displayed uncertainty aversion in the sense above is equivalent to the sum of two effects: one is a "pure" increase in total weight which keeps likelihood ratios approximately constant, the other is a revision of likelihood ratios that keeps total weight approximately constant.[15] It is straightforward to construct examples in which it turns out that the principal was not pessimistic enough when she formulated beliefs, and she over-estimated the likelihood of "good" states and under-estimated the likelihood of "bad' states. Obviously then a decrease in uncertainty aversion could be associated with a sharp decrease in profits for the principal. In such examples the effect of likelihood ratios revision is dominant. But we are presently more interested in "pure" reductions of uncertainty aversion, and we would like to focus our attention on those. This provides motivation for the following definition.

Definition 2 *Given two convex capacities ν and μ defined on an algebra Σ of subsets of Ω we say that μ displays pure uncertainty aversion not lower than ν if for every $T \in \Sigma$*

$$\nu(T) \geq \mu(T). \tag{9}$$

We say that μ displays pure uncertainty aversion higher than ν if at least one of the (9) holds with strict inequality.

It is clear that definition 2 is strictly stronger than DW's, and that a pure reduction in uncertainty aversion, while allowing revisions of likelihood ratios, will not allow drastic revisions.

As for case 2, in [6, section 3.2] we discussed what happens to the belief function ν_f induced by act f when the perceived state space becomes finer, and we showed that under certain assumptions on the perception correspondence, if we label ν'_f the distribution *after* the change in the state space, we have, for every $X \subseteq \mathcal{X}$

$$\nu'_f(X) \geq \nu_f(X)$$

which is just equation (9) applied to the distribution on $\mathcal{X}$ induced by f. Thus in both cases we obtain the following measure of comparative non-additivity.

Definition 3 *For two convex capacities μ and ν, defined on all subsets of a finite set $\mathcal{X}$, we say that ν is at least as purely non-additive as μ if (9) holds for every $T \subseteq \mathcal{X}$, and more purely non-additive than μ if moreover (9) holds with strict inequality for some T.*

Given definition 3 we can discuss the comparative statics of these types of changes in information. In what follows we compare solutions to the principal's problem before the change, when beliefs induced by $f \in \mathcal{F}$ are represented by ν_f, to solutions after the change, when beliefs are represented by ν'_f. We have the following immediate results.

[15] The "approximately" is necessary because we want to preserve convexity, so that some times exact changes will not be allowed. Hence differently from the standard example in consumer theory the decomposition in the two effects will not in general be unique.

Theorem 4 *For $f \in \mathcal{F}$, suppose that ν_f is at least as purely non-additive as ν'_f and moreover for every $T \subseteq X$ and every $g \in \mathcal{F}$*

$$\nu'_f(T) - \nu_f(T) \geq \nu'_g(T) - \nu_g(T) \tag{10}$$

then $Y_f \subseteq Y'_f$, that is, the set of incentive schemes that implement f is larger in the primed problem. Thus the principal's expected profit from implementing f in the primed problem will be not lower than that in the unprimed problem. That is,

$$\Psi'(f) \geq \Psi(f). \tag{11}$$

Proof: Let $Y \in Y_f$. It is clear from definition 3 that Y will satisfy the new individual rationality constraint, in fact notice that, if you let $z_i = u(y_i),\ i = 1, \ldots, n$,

$$\begin{aligned}\sum_{i=1}^{n}(z_i - z_{i+1})\,\nu'_f(\{x_1, \ldots, x_i\}) + z_n &\geq \sum_{i=1}^{n}(z_i - z_{i+1})\,\nu_f(\{x_1, \ldots, x_i\}) + z_n \\ &\geq \underline{u} + c(f).\end{aligned}$$

Moreover (10) implies that

$$\begin{aligned}\int_{\mathcal{X}} z(x)\,d\nu'_f(x) - \int_{\mathcal{X}} z(x)\,d\nu'_g(x) &\geq \int_{\mathcal{X}} z(x)\,d\nu_f(x) - \int_{\mathcal{X}} z(x)\,d\nu_g(x) \\ &\geq c(f) - c(g),\end{aligned}$$

so that also the (IC) are satisfied. Hence $Y \in Y'_f$, as required. The second statement follows by noticing that the Choquet integral of any function with respect to ν'_f is less than that of the same function with respect to ν_f. □

Condition (10) has the simple interpretation that the pure reduction in non-additivity for the capacity induced by action f should not be lower than the analogous reductions for the capacities induced by other actions. If this was not the case then it would be possible that the suddenly higher quality of information regarding another action g induces a violation of the incentive compatibility constraints. In the *ceteris paribus* case in which the change of information affects only one action (10) can obviously be dispensed with, and we can also say something about Y_g.

Corollary 5 *For $f \in \mathcal{F}$, suppose that ν_f is at least as purely non-additive as ν'_f and $\nu_g = \nu'_g$ for every $g \in \mathcal{F} \setminus \{f\}$, then $Y_f \subseteq Y'_f$ and $Y_g \supseteq Y'_g$. Consequently $\Psi'(f) \geq \Psi(f)$ and $\Psi'(g) \leq \Psi(g)$.*

Remark 3 Clearly the conditions of theorem 4 are sufficient, but not necessary. In particular we could substantially weaken them by letting them depend on the incentive scheme the principal is employing in the unprimed problem. That is, suppose that incentive scheme Y is optimal for implementing f. Then it is easy to find[16] a sufficient set of conditions, for Y

[16]See for instance the statement of theorem 6.

to be implementable in the primed problem, that only requires the probabilities associated with two monotone classes of subsets of $\mathcal{X}$ to be non-decreased.

The following result states sufficient conditions for the gain in expected profits to be strictly positive.

Theorem 6 *For every action $f \in \mathcal{F}$, $f \neq l$, and every point $Y \in Y_f$, it is possible to find $Y' \in Y'_f$ such that $Y > Y'$ if equation (10) holds for every $g \in \mathcal{F}$ and the following is true: (*) if π is the permutation of $\{1, \ldots, n\}$ such that $y_{1,\pi} \geq y_{2,\pi} \geq \cdots \geq y_{n,\pi}$, and $\{T_i\}_{i=1}^n$ is the monotone class of sets such that $T_i = \{x_{1,\pi}, \ldots, x_{i,\pi}\}$, then $\nu'_f(T_i) > \nu_f(T_i)$ for some $i = 1, \ldots, n$. In such a case the inequality in (11) will hold stricty.*

Proof: Let $Y \in Y_f$. By theorem 4, $Y \in Y'_f$ as well. Since $\nu'_f(T_i) > \nu_f(T_i)$ for some i, there is an $\epsilon > 0$ small enough so that

$$\int_{\mathcal{X}} z(x)\,\nu_f(x) - \epsilon - c(f) > \underline{u}$$

where as usual $z_i = u(y_i)$. Let $Z' = Z - \epsilon$, then obviously Z' satisfies the (IR)' and (IC)' constraints. Hence $Y' = u^{-1}(Z') \in Y'_f$ and $Y > Y'$ by monotonicity of u. □

Remark 4 For the *ceteris paribus* case, a result analogous to corollary 5 can be obtained. Clearly it is not generally possible to conclude that $\Psi'(g)$ is strictly lower than $\Psi(g)$, as the incentive compatibility constraint with respect to f is not necessarily binding at the optimal incentive scheme for implementing g.

The reader will certainly recognize the strong connection between this result and theorem 3. Basically it says that as (relative) pure non-additivity decreases in the sense discussed above the agent will be willing to perform action f (where f is *not* the least cost action) for a uniformly lower incentive scheme. Loosely speaking the reason is that his pessimism is constrained to operate on a smaller scale.

So far we have said nothing on the principal's choice of second best action and her profits. To do this let us start from the case in which the beliefs about all actions but one, f^* (which is not the least cost action), are unaffected by change, and ν'_{f^*} is less purely non-additive than ν_{f^*}. There are two possibilities: either f^* was the second best action or not. In the first case we have the following obvious result.

Corollary 7 *Suppose that $f^* \neq l$ is second best before change and that beliefs change as described in the paragraph above, then principal's expected profits will not decrease. In particular they will increase if hypothesis (*) of theorem 6 holds.*

If f^* was not second best before the change then nothing general can be said. While it is true that the profit from implementing f^* will increase, it does not have to be the case that it rises above the profit that the principal obtained before with some action g (that became more costly because of the tightening of the incentive compatibility constraint). This is of course due to the common prior assumption, which implies that any change in information will be shared by both players.

But the case in which information is received only on the implemented action is not as implausible as it might seem. It is conceivable that in some repeated agency situations new information about an action can *only* be gained by performing it. In such cases the discussion above implies that the principal will benefit from it. Also this would have as a result an extreme inertia in the principal's behavior. Even if some other action is judged to be better for the principal by some (perfectly informed) observer, the principal is doomed to get stuck on the action she tries first. Obviously this is a very extreme and rough picture, and it is clear that the development of a theory of the dynamics of information accrual will be crucial for obtaining more formal conclusions.

Finally we have to discuss the situation in which the change in information affects *all* actions, i.e., all distributions become less purely non-additive. As will be clear from our previous discussion it might be that the principal is better off[17] but she might as well be worse off. In general an obvious sufficient condition for her to be (weakly) better off is the following: suppose that f^* is the second best action before the change, and the principal implements it with the second best incentive scheme Y^*, then $Y^* \in Y'_{f^*}$, that is, Y^* implements f^* also after the change.

4.2 Informativeness in the Blackwell sense

The notion of informativeness which is better known in the agency literature is the one of Blackwell [2]. In fact Grossman and Hart proved [10, propositions 13 and 14] that in a problem with a revenue structure more informative in the Blackwell sense the principal will have profits not lower. A discussion of this requires a suitable generalization of the concept of informativeness in the sense of Blackwell to the non-additive case. To introduce it we first present a result of Gilboa and Schmeidler [8, theorems 3.3 and 4.3], which shows that every capacity is equivalent to an additive measure on a larger state-space. Let $\Sigma = 2^\Omega \setminus \emptyset$, i.e., the set of all non-empty subsets of Ω. Let u_T be the capacity defined as follows

$$u_T(A) = \begin{cases} 1 & A \supseteq T, \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 8 *For every capacity ν defined on a finite state-space Ω there are unique coefficients $\{\alpha_\nu(T)\}_{T\in\Sigma}$ such that*

$$\nu = \sum_{T\in\Sigma} \alpha_\nu(T) u_T. \tag{12}$$

Also for every action $f : \Omega \to \mathbf{R}$ one has

$$\int_\Omega f(\omega)\, d\nu(\omega) = \sum_{T\in\Sigma} \alpha_\nu(T)[\min_{\omega\in T} f(\omega)]. \tag{13}$$

[17] Generalizing corollary 7 we can obtain that as long as the change of beliefs on the second best action is uniformly larger (as per equation (10)) the principal will be better off.

Remark 5 The result can be (almost completely) generalized to an infinite state space [9]. In the finite case the formula for the coefficient $\alpha_\nu(T)$, $T \in \Sigma$, is the following:

$$\alpha_\nu(T) = \nu(T) \; - \sum_{\substack{I \subseteq \{1,\dots,m\} \\ I \neq \emptyset}} (-1)^{|I|+1} \nu(\cap_{i \in I} T_i) \tag{14}$$

where $T_i = T \setminus \{\omega_i\}$ and $T = \{\omega_1, \dots, \omega_m\}$. Equation (12) can be equivalently written as follows: for every $A \subseteq \Omega$,

$$\nu(A) = \sum_{T \subseteq A} \alpha_\nu(T).$$

The $\alpha_\nu(T)$ will be non-negative for every $T \in \Sigma$ if and only if ν is a belief function. Also $\alpha_\nu(T) = 0$ for every $T \in \Sigma$ such that $|T| > 1$ if and only if ν is additive. α_ν is called the *Möbius transform* of ν.

Given this it is natural to generalize Blackwell's definition to capacities by requiring that it holds on their Möbius transforms. Therefore let $\Sigma = 2^X \setminus \emptyset$ and let m be its cardinality (obviously $m = 2^n - 1$), so that α can be written as a vector in $\mathbf{R}^m$. In particular we assume that the first n entries correspond to $\alpha(\{x_1\}), \dots, \alpha(\{x_n\})$.

Definition 4 *Let ν and μ be two different capacities. Then we say that ν is more informative (or sufficient) in the (generalized) Blackwell sense than μ if there is a $m \times m$ Markov matrix*[18] *Q such that*

$$\alpha_\mu = Q^T \alpha_\nu.$$

As it turns out, this definition is too general. Without having any constraints on Q basically everything is possible, in the sense that we allow the transfer of weight from singleton to non-singeton sets and *vice versa*. Thus it is possible that ν is extremely non-additive (giving non-zero probability only to Ω) and μ is additive. Clearly in a situation like this it would be impossible to prove that the principal is better off with beliefs like ν than with μ (by arguments similar to the ones presented in section 4.1), so that Grossman and Hart's result would not hold. It seems natural that a proper generalization of Blackwell's definition ought to bar the possibility of weight transfer between sets of different cardinality, or at least between singletons and all other sets, by requiring Q to have a block diagonal structure. In this way we would be able to isolate the statistical effect of information dispersion from changes in non-additivity. The following example shows that even this is not enough to generalize Grossman and Hart's result.

Example 3 There are only two actions, f and l, and two outcomes, i.e., $\mathcal{X} = \{x_1, x_2\}$. Agent's utility function is $u(y) = y^{1/2}$, $\underline{u} = 0$, and $c(f) = 4.25$, $c(l) = 4$. We assume that for every action $a \in \{f, l\}$

$$\alpha_{\nu'_a} = Q^T \alpha_{\nu_a}$$

[18] A Markov matrix is a matrix $Q = [q_{ij}]$ such that $q_{ij} \geq 0$ for every i, j and $\sum_j q_{ij} = 1$ for every i.

where Q is the following matrix

$$\begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Notice that there is no transfer of weight from singleton sets to $\mathcal{X}$, the only non-singleton, and that Q does not have a left inverse which is a Markov matrix, so that ν'_a is not more informative than ν_a. In the unprimed problem $\mathcal{X} = \{30, 6\}$ and beliefs are given by

$$\begin{aligned} \nu_{f1} &= 3/8 & \nu_{f2} &= 3/8 \\ \nu_{l1} &= 1/4 & \nu_{l2} &= 5/16. \end{aligned}$$

Thus in the primed problem[19] $\mathcal{X}' = \{24, 12\}$ and beliefs are

$$\begin{aligned} \nu'_{f1} &= 1/2 & \nu'_{f2} &= 1/4 \\ \nu'_{l1} &= 3/8 & \nu'_{l2} &= 3/16. \end{aligned}$$

Under the conditions stated, if beliefs are additive Grossman and Hart's result implies that the principal's expected profit is strictly lower in the primed problem. It is a straightforward exercise to check that in the unprimed problem the principal will be indifferent between optimally implementing f with the incentive scheme $(30.25, 12.25)$ and implementing l with the flat 16 incentive scheme, for an expected profit of -4.

On the other hand in the primed problem the principal can optimally implement f with the incentive scheme $(27.5625, 10.5625)$, which gives him an expected profit of -2.3125, but she chooses l as second best action for an expected profit of 1/2. Thus the principal is definitely better off in the (less informative) primed problem than in the unprimed problem.

In the simple two outcome case we have studied conditions on Q which insure the validity of Grossman and Hart's result,[20] but they were scarcely justifiable, and involved weight transfer towards singleton sets. Also it was unclear how they would generalize to any finite number of outcomes. Thus we conclude that, unless some different generalization of Blackwell informativeness is devised, in general it seems not possible to extend Grossman and Hart's result in a palatable way.

References

[1] Dimitri P. Bertsekas. Necessary and sufficient conditions for existence of an optimal portfolio. *Journal of Economic Theory*, 8:235–247, 1974.

[19] If $\mathbf{x}$ is the vector of outcomes in the unprimed problem, the vector $\mathbf{x}'$ in the primed problem satisfies $\mathbf{x}' = Q^T\mathbf{x}$. Obviously the non-separability of the principal's objective function imposes to discuss principal's profits as a whole so that also this assumption has to be made (see Grossman and Hart [10, proposition 14]).

[20] This works by mimicking their proof style. Obviously there could be a different proof strategy which yielded the same result with more appealing conditions, but we could not find any.

[2] David Blackwell. Comparison of experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951.

[3] Gustave Choquet. Theory of capacities. *Annales de l'Institut Fourier (Grenoble)*, 5:131–295, 1953.

[4] James Dow and Sergio Werlang. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, 60:197–204, 1992.

[5] Daniel Ellsberg. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 75:643–669, 1961.

[6] Paolo Ghirardato. Coping with ignorance: Unforeseen contingencies and non-additive uncertainty. *Economic Theory*, 2001. forthcoming; `http://masada.hss.caltech.edu/~paolo/uco-IVwp.pdf`.

[7] Itzhak Gilboa. Expected utility with purely subjective non-additive probabilities. *Journal of Mathematical Economics*, 16:65–88, 1987.

[8] Itzhak Gilboa and David Schmeidler. Additive representations of non-additive measures and the Choquet integral. *Annals of Operations Research*, 52:43–65, 1994.

[9] Itzhak Gilboa and David Schmeidler. Canonical representation of set functions. *Mathematics of Operations Research*, 20:197–212, 1995.

[10] Sanford J. Grossman and Oliver D. Hart. An analysis of the principal-agent problem. *Econometrica*, 51:7–45, 1983.

[11] Ebbe Hendon, Hans Jørgen Jacobsen, Birgitte Sloth, and Torben Tranæs. Expected utility with lower probabilities. *Journal of Risk and Uncertainty*, 8:197–216, 1994.

[12] Bengt Holmström. Moral hazard and observability. *Bell Journal of Economics*, 10:74–91, 1979.

[13] Jean-Yves Jaffray and Peter P. Wakker. Decision making with belief functions: Compatibility and incompatibility with the sure-thing principle. *Journal of Risk and Uncertainty*, 8:255–271, 1994.

[14] Frank H. Knight. *Risk, Uncertainty and Profit.* Houghton Mifflin, Boston, 1921.

[15] David M. Kreps. *A Course in Microeconomics.* Princeton University Press, Princeton, 1990.

[16] Sujoy Mukerji. Understanding the non-additive probability decision model. *Economic Theory*, 9:23–46, 1997.

[17] David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587, 1989.

[18] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey, 1976.

[19] Peter P. Wakker. *Additive Representations of Preferences*. Kluwer, Dordrecht, 1989.