Hybrid Decision Model and the Ranking of Experiments*

Edi Karni† Zvi Safra‡

December 21, 2021

Abstract

Departing from the reduction of compound lotteries axiom on multi-stage lotteries, this paper proposes a new hybrid model to analyze decision trees. Applied to multi-stage decision trees induced by experiments, Blackwell's (1953) definition of the relation "more informative" on the set of information structures is equivalent to experiments being more valuable to a class of non-expected utility preferences. This result extends Blackwell's theorem and provides new insights regarding the evaluation of information produced by experiments.

Keyword: Blackwell's theorem; comparison of experiments; reduction of compound lotteries; value of information; hybrid decision analysis.

JEL classification: D81, D83

1 Introduction

From a decision making point of view, experimentation is valuable because it provides information that helps decision makers choose courses of actions whose payoffs are higher in the states that are more likely to obtain. Blackwell (1953) formalized this perception as a binary relation: "more informative than" on the set of experiments. According to Blackwell, one experiment is more informative than another if, *for every set of feasible actions, it yields a richer menu of experiment-wise expected payoffs* (i.e., expected-loss vectors) each of which corresponds to an action taken contingent on the experimental observations. Blackwell characterized this relation by proving that one experiment is more informative than another if and only if the information content of the latter is obtained by garbling the information content of the former. Equivalently, an experiment is more informative if it allows choices that have higher expected utility. We refer to this equivalence as Blackwell's theorem.[1]

*Useful comments by Edward Schlee and an anonymous refree are gratefully acknowledged.

†Johns Hopkins University, Department of Economics, e-mail: karni@jhu.edu

‡Warwick Business School, University of Warwick, e-mail: zvi.safra@wbs.ac.uk

[1] Cremer (1982), Leshno and Spector (1992), and de Oliviera (2018) provide simple proofs of Blackwell's theorem.

Seen in this way, being better informed seems unambiguously beneficial. Thus, the equivalence between ranking experiments by their information content and their ranking by the expected utility criterion seems oddly restrictive. This equivalence is particularly disconcerting in view of experimental evidence suggesting that subjects systematically violate the tenets of expected utility theory – the sure thing principle and the independence axiom – and the proliferation, over the last 40 years, of non-expected utility models of decision making under risk and under uncertainty.

To grasp the issue, consider a decision maker facing a choice among feasible actions whose consequences depend on the realization of some underlying states. Suppose that the likelihood of the various states materializing is quantified by a (prior) probability distribution function. Before choosing an action, the decision maker receives a signal (i.e., an observation), produced by an experiment, that informs him about the likely realization of the states. Upon receiving such signal, the decision maker invokes Bayes' rule to update the prior state probabilities and then proceeds to choose an action from the feasible set. This process may be thought of as two-stage compound lottery. In the first stage, the experiment produces a signal, according to some probability distribution on the set of signals, following which the decision maker chooses an action. In the second stage, a state is selected (according to the posterior distribution) and the decision maker is awarded the prize that corresponds to the image of the selected state under the chosen action.[2] The question is how decision makers perceive this two-stage lottery.

We argue that the critical aspect of the expected utility model that underlies Blackwell's theorem is the way these compound lotteries are handled. The standard way of handling compound lotteries is to apply the reduction of compound lotteries axiom (henceforth, RCLA). This axiom asserts that a multi-stage lottery is reduced to a single-stage lottery by attributing to each ultimate payoff a probability equal to the product of the probabilities on the events that lead to it. In the context of the choice of experiments, the RCLA assigns the ultimate outcome the probability of the signal multiplied by the posterior probabilities of the states to which the chosen course of action assigns that outcome.

Analysis that treats the two-stage process as equivalent to its one-stage reduction runs the risk of ignoring subtleties that beset the extensive form decision process. Wakker (1988) and Safra and Sulganik (1995) showed that departing from the independence axiom and maintaining the RCLA implies a widespread and robust aversion towards information.

An alternative procedure of handling compound lotteries is to use the certainty-equivalent reduction. In this procedure, multi-stage lotteries are reduced to single-stage lotteries by folding back the lottery tree, replacing the lotteries along the branches by their certainty equivalents. Schlee (1990) and Safra and Sulganik (1995) showed that in non-expected utility theories, this procedure implies that information is not always valuable.

We propose a hybrid decision model that invokes the RCLA in the second stage of

[2] If the prize itself is a lottery ticket then the procedure described above amounts to three-stage lottery in which, in the third and final stage, the lottery corresponding to the image of the selected state under the chosen action is played out to determine the prizes.

the decision-making process and a procedure analogous to certainty-equivalent reduction in the first stage. Application of the RCLA, which seems compelling when the transition between the stages is automatic, seems less so when the two stages are separated by an intermediate decision.

To formalize this idea, we propose a new model, dubbed the *hybrid model*, and show that it identifies a class of preferences that unambiguously value one experiment over another if and only if the information content of the latter is obtained by garbling that of the former. The expected utility model is a special case of this class. In fact, it is the only hybrid model that is consistent with the RCLA. Application of the hybrid model provides new insight into the manner in which information is evaluated. While admitting the possibility that some decision makers may not consider information to be unambiguously beneficial, it maintains that, in Blackwell's analytical framework, such behavior is unreasonable.

The surprising (difficult) aspect of Blackwell's theorem is that a more informative experiment (that is, one that affords better decisions by the expected utility criterion) implies clearer signals. In this paper, informativeness corresponds to an experiment being more valuable in the sense of affording better decisions for a broader set of preferences, including expected utility preferences. Consequently, this direction of the proof relies on Blackwell's theorem. The novelty of this paper is the observation that the full power of expected utility – in particular, the RCLA – is not needed for Blackwell's result. To the best of our knowledge, ours is the only nontrivial model in which a large class of non-expected utility preferences can satisfy this direction of Blackwell's theorem.[3]

The rest of the paper is organized as follows. The next section provides a brief review of Blackwell's (1953) theorem. Section 3 reviews the reduction procedures. Section 4 introduces and characterizes the hybrid decision model. Section 5 extends Blackwell's theorem. Section 6 discusses the value of information and reviews the related literature.

2 The Analytical Framework and Blackwell's Theorem

2.1 The analytical framework

Let $S = \{s_1, ..., s_n\}$ be a finite set of *states*, and denote by $\Delta(S)$ the simplex in $\mathbb{R}^n$. Let X be a set of *outcomes*, and denote by $\Delta(X)$ the set of simple probability distributions on X referred to as *lotteries*.[4] For all $p, q \in \Delta(X)$ and $\alpha \in [0, 1]$ define $\alpha p + (1 - \alpha) q \in \Delta(X)$ by $(\alpha p + (1 - \alpha) q)(x) = \alpha p(x) + (1 - \alpha) q(x)$, for all $x \in X$. Mappings from S to $\Delta(X)$ are referred to as *acts*. Acts represent potential courses of action. The set of all acts is denoted by $\mathcal{H}$. For all $f, g \in \mathcal{H}$ and $\alpha \in [0, 1]$ define $\alpha f + (1 - \alpha) g \in \mathcal{H}$ by $(\alpha f + (1 - \alpha) g)(s) = \alpha f(s) + (1 - \alpha) g(s)$, for all $s \in S$. Constant acts (i.e., acts that assign the same image to every state) are identified with elements of $\Delta(X)$. Thus, $\Delta(X) \subset \mathcal{H}$. Throughout, we

[3] Li and Zhou (2016) carried out an analysis with commitment, in which decision makers are not allowed to change their choices when information is revealed (and in the context of ambiguity).

[4] Simple probability distributions are probability distribution functions with finite supports.

denote by δ_x the distribution function that assigns $x \in X$ the unit probability mass and $\Delta(\Delta(X))$ denotes the set of simple probability distributions with supports in $\Delta(X)$.

Let $\mathbb{Y}$ be a finite set whose generic element, y, is a *signal.* An *experiment,* $\widetilde{y}$, also called an *information structure,* is identified with a conditional probability distribution μ on $\mathbb{Y} \times S$ such that, for all $s \in S$, $\mu(y \mid s)$ is the conditional probability of $y \in \mathbb{Y}$ given s. Let $\mathcal{Y}$ denote the set of all experiments. The *image* of $\tilde{y} \in \mathcal{Y}$ is the set $Y = \{y \in \mathbb{Y} \mid \exists s \in S \ : \mu(y \mid s) > 0\}$ of all potentially realized signals under $\tilde{y}$. The experiment $\widetilde{y}$ can be identified with the $|S| \times |Y|$ right-stochastic matrix, $P(\tilde{y})$, whose generic element is $\mu(y \mid s)$. For a given prior $\pi \in \Delta(S)$, denote $\mu(y) = \Sigma_{s\in S}\mu(y \mid s)\pi(s)$.

2.2 Blackwell's theorem

Consider an expected utility-maximizing decision maker characterized by a utility function u in $\mathcal{U}$, the set of all real-valued functions on X. Consider a probability distribution $\pi \in \Delta(S)$ and a non-empty, compact set $B \subseteq \mathcal{H}$ of feasible acts. Assume that π is strictly positive (i.e., $\pi(s) > 0$ for all $s \in S$) and that, before choosing an act from B, the decision maker observes the signal y, generated by an experiment $\tilde{y} \in \mathcal{Y}$, and updates the prior distribution π according to Bayes' rule to obtain the posterior probability distribution $\pi(s \mid y) = \mu(y \mid s)\pi(s)/\mu(y)$, for all $s \in S$. The decision maker's ex post problem is:

$$\max_{f\in B} \Sigma_{s\in S}\pi(s \mid y)\Sigma_{x\in X}u(x) f(s)(x). \tag{1}$$

Denoting the maximal value by $U(u, \pi(\cdot \mid y), B)$, the expected utility associated with the experiment $\tilde{y}$ is given by:

$$\hat{U}(\tilde{y}; u, \pi, B) := \Sigma_{y\in Y}\mu(y) U(u, \pi(\cdot \mid y), B). \tag{2}$$

Definition 1: An experiment $\tilde{y}$ is *more informative* than another experiment $\tilde{y}'$ if, for all $(u, B) \in \mathcal{U} \times 2^{\mathcal{H}}\backslash\varnothing$,

$$\hat{U}(\tilde{y}; u, \pi, B) \geq \hat{U}\left(\tilde{y}'; u, \pi, B\right).$$

Let $\mathcal{M}$ be the set of $|Y| \times |Y'|$ right-stochastic matrices dubbed *garbling matrices.* If there exists a garbling matrix $M \in \mathcal{M}$ such that $P(\tilde{y}) M = P'(\tilde{y}')$, then M introduces noise that blurs the information in $P(\tilde{y})$. That is, for each entry $\mu'(y'_r \mid s)$ of $P'(\tilde{y}')$, $\mu'(y'_r \mid s) = \Sigma_{y_j\in Y}\mu(y_j \mid s) m_{jr}$ (where $\Sigma_{\{r|y_r\in Y'\}}m_{jr} = 1$ for all $\{j \mid y_j \in Y\}$).

Definition 2: An experiment $\tilde{y}$ is *sufficient* for $\tilde{y}'$ if the corresponding information structures satisfy $P(\tilde{y}) M = P'(\tilde{y}')$, for some $M \in \mathcal{M}$.

With these definitions in mind Blackwell's theorem is stated as follows:

Blackwell's Theorem: *An experiment $\tilde{y}$ is more informative than another experiment $\tilde{y}'$ if and only if $\tilde{y}$ is sufficient for $\tilde{y}'$.*

That is, the relation "being sufficient" is equivalent to being ranked higher by all preference relations that admit expected utility representations.

3 Informative Signals and Reduction Procedures

3.1 Signals

According to Blackwell's theorem more, informative experiments produce clearer signals in the following sense: When comparing two experiments, every posterior distribution produced by the less informative experiment is a weighted average of the posterior distributions produced by the more informative experiment (i.e., the sufficient experiment). Formally, let $\tilde{y}$,$\tilde{y}' \in \mathcal{Y}$ with images Y and Y', respectively. Suppose that $\tilde{y}$ is sufficient for $\tilde{y}'$, hence there exists $M \in \mathcal{M}$ such that $P(\tilde{y}) M = P'(\tilde{y}')$. Observe that the equalities

$$\mu'\left(y'_r \mid s\right) = \Sigma_{\{j|y_j \in Y\}} \mu\left(y_j \mid s\right) m_{jr}, \tag{3}$$

for all $y'_r \in Y'$, $y_j \in Y$ and $s \in S$, imply

$$\mu'\left(y'_r\right) = \Sigma_{\{j|y_j \in Y\}} \mu\left(y_j\right) m_{jr}. \tag{4}$$

Therefore, by Bayes' rule,

$$\begin{aligned}
\pi\left(s \mid y'_r\right) &= \frac{\pi(s)\,\mu'\left(y'_r \mid s\right)}{\Sigma_{s' \in S}\pi(s')\,\mu'\left(y'_r \mid s'\right)} = \frac{\pi(s)\,\Sigma_{\{j|y_j \in Y\}}\mu\left(y_j \mid s\right) m_{jr}}{\mu'\left(y'_r\right)} \\
&= \frac{\Sigma_{\{j|y_j \in Y\}}\pi(s)\,\mu\left(y_j \mid s\right) m_{jr}}{\mu'\left(y'_r\right)} = \frac{\Sigma_{\{j|y_j \in Y\}}\frac{\mu(y_j)}{\mu(y_j)}\pi(s)\,\mu\left(y_j \mid s\right) m_{jr}}{\mu'\left(y'_r\right)} \\
&= \frac{\Sigma_{\{j|y_j \in Y\}}\mu\left(y_j\right)\pi(s \mid y_j) m_{jr}}{\mu'\left(y'_r\right)} = \Sigma_{\{j|y_j \in Y\}}\frac{\mu\left(y_j\right) m_{jr}}{\mu'\left(y'_r\right)}\pi\left(s \mid y_j\right).
\end{aligned} \tag{5}$$

and the posterior distributions satisfy,

$$\pi\left(\cdot \mid y'_r\right) = \Sigma_{\{j|y_j \in Y\}}\frac{\mu\left(y_j\right) m_{jr}}{\mu'\left(y'_r\right)}\pi\left(\cdot \mid y_j\right).$$

Consequently, for each act-posterior probability pair $(f, \pi(\cdot \mid y'_r)) \in \mathcal{H} \times \Delta(S)$ that is feasible under the less informative experiment, corresponds a set $\{(f, \pi(\cdot \mid y_j)) \mid y_j \in Y\} \subset \mathcal{H} \times \Delta(S)$ of act-posterior probability pairs of the more informative experiment. Thus, from the ex ante viewpoint, the more informative experiment offers a richer set of opportunities to match feasible acts to the perceived likelihood of the states depicted by their posterior probabilities.

3.2 Reduction procedures

Consider a pair $(f,\pi) \in \mathcal{H} \times \Delta(S)$ of an act and a probability distribution on S. The pair may be regarded as a two-stage lottery in $\Delta(\Delta(X))$ in which, in the first stage, a state $s \in S$ is drawn at random according to the distribution π and, in the second stage, an outcome $x \in X$ is determined by the lottery $f(s) \in \Delta(X)$.

The most common way to reduce such a two-stage lottery to a one-stage lottery in $\Delta(X)$ is the *reduction of compound lotteries axiom* (RCLA). This reduction procedure identifies (f,π) with the one-stage mixture lottery $\Sigma_{s\in S}\pi(s) f(s) \in \Delta(X)$. The axiomatic structure underlying expected utility theory implicitly utilizes RCLA.

Another way to reduce two-stage lotteries is the *certainty-equivalent reduction.*[5] This reduction procedure assumes that the decision maker possesses a preference relation over $\Delta(X)$ and that, for every lottery $f(s) \in \Delta(X)$, there exists an element $c(f(s)) \in X$, the *certainty equivalent* of $f(s)$, such that the decision maker is indifferent between $f(s)$ and $\delta_{c(f(s))}$.[6] The certainty-equivalent reduction identifies (f,π) with the one-stage mixture lottery $\Sigma_{s\in S}\pi(s)\,\delta_{c(f(s))} \in \Delta(X)$. Like RCLA, this reduction is implicit in expected utility theory.

In general the resulting one-stage lotteries $\Sigma_{s\in S}\pi(s) f(s)$ and $\Sigma_{s\in S}\pi(s)\,\delta_{c(f(s))}$ are not equivalent. However, under expected utility they, are closely related in the sense that the decision maker is always indifferent between them.

To see how these reductions affect the evaluation of experiments, let a probability distribution $\pi \in \Delta(S)$ and a nonempty set $B \subseteq \mathcal{H}$ of feasible acts be given, and consider a decision maker facing an experiment $\tilde{y} \in \mathcal{Y}$. From the decision maker's viewpoint, this decision problem requires choosing acts in B contingent on the realization of signals produced by $\tilde{y}$. This problem can be described as a three-stage compound lottery. In the first stage, a signal $y \in Y$ is drawn according to the distribution μ. Contingent on the signal and the corresponding posterior distribution $\pi(\cdot \mid y)$, an act $f^*(\pi(\cdot \mid y)) \in B$ is chosen. In the second stage, a state s is selected according to the posterior distribution $\pi(\cdot \mid y)$, and the lottery $f^*(\pi(\cdot \mid y))(s) \in \Delta(X)$ is awarded as a prize. In the third and final stage, the lottery $f^*(\pi(\cdot \mid y))(s)$ determines the final outcome $x \in X$. Using the previous notation, for each y, the pair $(f^*(\pi(\cdot \mid y)), \pi(\cdot \mid y)) \in \mathcal{H} \times \Delta(S)$ describes the final two stages.

It is worth underscoring that the sequence of events described above is depicted, as it should be, from the point of view of the decision maker who receives the signal and must act upon the information it provides. The problem may be reformulated so that the state is drawn first according to the prior $\pi \in \Delta(S)$, followed by a signal drawn from $\mathbb{Y}$ according to the conditional distribution $\mu(\cdot \mid s)$. Because the state is not observable, however, the decision maker may not be able to distinguish between nodes resulting from different states that potentially yield the signal (while he must act solely on the basis of the information produced by the signal). Consequently, from the decision maker's point

[5] For more detailed discusssion and application of certainty equivalent reduction, see Segal (1987).
[6] The existence of certainty equivalents depends on the richness of the set of outcomes X.

of view, reformulating the problem so that the state is selected in the first stage followed by the draw of a signal is substantively equivalent to the formulation above but would complicate the notations and clutter the exposition.

Applying the RCLA to the second stage reduces the pairs $(f^*(\pi(\cdot \mid y)), \pi(\cdot \mid y))$ to the one-stage lotteries $\Sigma_{s \in S} \pi(s \mid y) f^*(\pi(\cdot \mid y))(s) \in \Delta(X)$. Applying the RCLA again, this time to the first stage, further simplifies the three-stage compound lottery and reduces it to the one-stage mixture lottery

$$\sum_{y \in Y} \mu(y) \sum_{s \in S} \pi(s \mid y) f^*(\pi(\cdot \mid y))(s). \tag{6}$$

The process under which RCLA is applied twice is called the RCLA *procedure*; it is commonly used, in conjunction with consequentialism, to analyze behavior of non-expected utility decision makers. The procedure implies a widespread and robust aversion to information. The reason for this aversion is that the RCLA, in conjunction with consequentialism, imply that preference relations display dynamic consistency (that is, the optimal continuation of the sequential choice process as of any decision node agrees with the optimal contingent plan, made at the outset of the process, for that node) if and only if it satisfies the independence axiom of expected utility theory.

To grasp this assertion, consider the following simple example. A decision maker faces a choice between a lottery A and an equal chance of receiving another lottery, B, or a choice between two lotteries, C and D (all lotteries are in $\Delta(X)$). At the initial decision node, the decision maker must choose between A and two contingent plans, one calling for the choice of C and the other the choice of D in case he finds himself at the second decision node. Under the RCLA, the former contingent plan induces the lottery $0.5B + 0.5C$, and the latter the lottery $0.5B + 0.5D$. If the preference relation does not satisfy the independence axiom, there are always lotteries C and D such that $C \succ D$ and $0.5B + 0.5D \succ 0.5B + 0.5C$. Suppose further that, $0.5B + 0.5D \succ A \succ 0.5B + 0.5C$. Then the decision maker rejects lottery A in favor of pursuing the contingent plan that call for the choice of D in the second decision node. However, if he happens to find himself at that node and is satisfying consequentialism, he will choose C, thereby displaying dynamic inconsistency. Furthermore, realizing this dynamic inconsistency, the decision maker would prefer A over the contingent plan that yields $0.5B + 0.5C$.[7] Consequently, in our case, evaluating the optimal acts $f^*(\pi(\cdot \mid y))$ from an *ex ante* point of view, by looking at the one-stage lottery (6), the decision maker who does not abide by the independence axiom may find out that he prefers to replace some of these *ex post* optimal acts by others. However, when he finds himself at the second stage, optimality dictates that he choose the acts $f^*(\pi(\cdot \mid y))$.

Certainty-equivalent reduction can be applied twice to yield the one-stage mixture lottery

$$\sum_{y \in Y} \mu(y) \, \delta_{c\left(\Sigma_{s \in S} \pi(s|y) \delta_{c(f^*(\pi(\cdot|y))(s))}\right)}. \tag{7}$$

[7] For a detailed analysis see Karni and Schmeidler (1991), and Karni and Safra (1989).

Under this certainty-equivalent reduction *procedure*, information is not always valuable for non-expected utility decision makers. Unlike the RCLA, however, the reason here is not dynamic inconsistency (it is easy to verify that, viewing the one-stage lottery (7) *ex ante*, the decision maker is always in agreement with his planned *ex post* optimal acts $f^*(\pi(\cdot \mid y))$). It is that the mixture and certainty-equivalent operators are not commutative for non-expected utility decision makers. Thus, having larger sets of options under a sufficient experiment does not necessarily translate to being in a preferable situation.

An alternative procedure, which we propose and study in this paper, applies the RCLA to the second stage and a reduction analogous to the certainty equivalent to the first stage.[8] Under this *hybrid procedure*, which is characterized in the next section, for each signal y, the second-stage reduction induces a set of one-stage lotteries $\{\Sigma_{s\in S}\pi(s \mid y) f(s) \mid f \in B\}$. Assuming that these lotteries are evaluated by some utility functional v (see next section), let f_y^* be a maximizer of v over this set (that is, $f_y^* \in \arg\max_{f\in B} v(\Sigma_{s\in S}\pi(s \mid y) f(s))$). Then the value of the experiment $\tilde{y}$ is given by

$$\sum_{y\in Y} \mu(y) v\left(f_y^*\right).$$

The justification for applying distinct procedures to the different stages is the nature of the uncertainties involved. In the second stage, given the act and the (updated) state probabilities, the outcome is selected "algorithmically" without interference by the decision maker. By contrast, after the first stage, corresponding to each signal there is an interim step at which the decision maker interferes by updating the state probabilities and choosing an act. This aspect of the dynamic process suggests that decision makers may regard the first stage as qualitatively distinct from the later stages and, consequently, treat them differently. Specifically, according to the hybrid procedure, assessing the value of experiments, decision makers envision the acts they would choose contingent on the signals, assign these acts utility values, and take the mean utility values as the value of the experiment.

4 Characterization of the Hybrid Representation

Assume that X is a compact topological space and consider a probability distribution $\pi \in \Delta(S)$, a nonempty compact set $B \subseteq \mathcal{H}$, and an experiment $\tilde{y} \in \mathcal{Y}$. As above, the experiment can be identified with a two-stage decision tree in which a signal $y \in Y$ is realized in the first stage and, in the second stage, at the decision node associated with y, the decision maker chooses an act $f \in B$ so as to maximize his ex post preferences. We assume that in the second stage the RCLA is applied. Formally, for all $f \in \mathcal{H}$, $f \sim \Sigma_{s\in S}\pi(s \mid y) f(s)$. Consequently, by choosing an act $f_y^* \in B$, the decision maker is faced with the reduced lottery $\Sigma_{s\in S}\pi(s \mid y) f_y^*(s) \in \Delta(X)$. From an ex ante viewpoint, the ex post choices are seen as the $|Y|+1$ tuple $(\mu, (f_y^*)_{y\in Y}) \in \Delta(\mathbb{Y}) \times \Delta(X)^{|\mathbb{Y}|}$, and the whole experiment can

[8] The analogous reduction does not require the existence of certainty equivalents.

be identified with a subset of all $|\mathbb{Y}|+1$ tuples $\left(\mu,\left(p_1,...,p_{|\mathbb{Y}|}\right)\right)\in\Delta(Y)\times\Delta(X)^{|\mathbb{Y}|}$. To analyze all possible experiments, assume that the cardinality of $\mathbb{Y}$ (the set of all signals) is N ($<\infty$), and consider the set of all $N+1$ tuples $(\mu,(p(y))_{y\in\mathbb{Y}})\in\Delta(\mathbb{Y})\times\Delta(X)^N$. A decision maker is characterized by two complete and transitive, preference relations: (i) an ex ante preference relation $\succcurlyeq$ on $\Delta(\mathbb{Y})\times\Delta(X)^N$, which ranks decision trees, and (ii) an ex post preference relation $\succcurlyeq^*$ on $\Delta(X)$, that is used to choose optimal acts at the realized decision nodes. Note that for the preference relation $\succcurlyeq$ to have the domain indicated above, it is implicitly assumed that the decision maker is able to express preferences over all conceivable information structures, so that every conceivable distribution on $\mathbb{Y}$ is attainable.

Continuity: (of $\succcurlyeq$ and $\succcurlyeq^*$ with respect to the corresponding Euclidean topology) The sets $\{(\mu',(p'(y))_{y\in\mathbb{Y}}\in\Delta(\mathbb{Y})\times\Delta(X)^N\mid(\mu,(p(y))_{y\in\mathbb{Y}}\succcurlyeq(\mu',(p'(y))_{y\in\mathbb{Y}})\}$ and $\{(\mu',(p'(y))_{y\in\mathbb{Y}}\in\Delta(\mathbb{Y})\times\Delta(X)^N\mid(\mu',(p'(y))_{y\in\mathbb{Y}}\succcurlyeq(\mu,(p(y))_{y\in\mathbb{Y}})\}$ are closed for all $\left(\mu,(p(y))_{y\in\mathbb{Y}}\right)\in\Delta(\mathbb{Y})\times\Delta(X)^N$, and the sets $\{p\in\Delta(X)\mid q\succcurlyeq^* p\}$, $\{p\in\Delta(X)\mid p\succcurlyeq^* q\}$ are closed for all $q\in\Delta(X)$.

Henceforth we assume that both $\succcurlyeq$ and $\succcurlyeq^*$ are continuous binary relations. To capture consequentialism, we assume that the ex post preference relation $\succcurlyeq^*$ is independent of the realized signal y and of the probability distribution μ.

The next axiom asserts that if a signal y is certain to obtain, then the ex ante ranking of any two lotteries agrees with their ex post ranking. Formally, define $\mathbf{p}=(p(y))_{y\in\mathbb{Y}}\in\Delta(X)^N$, then

Consistency: *For all* $\mathbf{p},\mathbf{q}\in\Delta(X)^N$ *and* $y\in\mathbb{Y}$, $(\delta_y,\mathbf{p})\succcurlyeq(\delta_y,\mathbf{q})$ *if and only if* $p(y)\succcurlyeq^* q(y)$.

The next axiom asserts that independence applies to the first stage of the decision-making process. To formalize this idea we introduce the partial mixture operation on $\Delta(\mathbb{Y})\times\Delta(X)^N$: For all $(\mu,\mathbf{p})$, $(\nu,\mathbf{p})\in\Delta(\mathbb{Y})\times\Delta(X)^N$ and $\lambda\in[0,1]$, define $\lambda(\mu,\mathbf{p})+(1-\lambda)(\nu,\mathbf{p})=(\lambda\mu+(1-\lambda)\nu,\mathbf{p})$. The mixture $\lambda\mu+(1-\lambda)\nu\in\Delta(\mathbb{Y})$ depicts a potential distribution on the signals space induced by an experiment. Alternatively, it can be interpreted as a choice between two experiments that is decided by a coin flip.

Consider a decision maker who is indifferent between the alternatives $(\mu,\mathbf{p})$ and $(\mu',\mathbf{q})$ and prefers the alternative $(\nu,\mathbf{p})$ over $(\nu',\mathbf{q})$. Suppose that, facing a choice between the decision trees $A=(\lambda\mu+(1-\lambda)\nu,\mathbf{p})$ and $B=(\lambda\mu'+(1-\lambda)\nu',\mathbf{q})$ he reasons that if the event whose probability is λ obtains and he has chosen A, he is faced with the alternative $(\mu,\mathbf{p})$; if he has chosen B, he faces the alternative $(\mu',\mathbf{q})$. Conditional on the realization of this event, he is indifferent between A and B. By the same logic, he would prefer A over

B conditional on the realization of the complementary event whose probability is $1-\lambda$. Consequently, he prefers A over B unconditionally.[9] Formally,

First-Stage Independence: *For all* $(\mu,\mathbf{p})$, $(\nu,\mathbf{p})$,$(\mu',\mathbf{q})$, $(\nu',\mathbf{q})$ *in* $\Delta(\mathbb{Y})\times\Delta(X)^N$ *and* $\lambda\in[0,1]$, *if* $(\mu,\mathbf{p})\sim(\mu',\mathbf{q})$ *then* $(\nu,\mathbf{p})\succcurlyeq(\nu',\mathbf{q})$ *if and only if* $(\lambda\mu+(1-\lambda)\nu,\mathbf{p})\succcurlyeq(\lambda\mu'+(1-\lambda)\nu',\mathbf{q})$.

The next theorem provides the hybrid representation. It shows that an ex ante preference relation satisfying consistency (with respect to an ex post preference relation) and first-stage independence is representable as a weighted sum of the ex post utilities.

Theorem 1: *Let* $\succcurlyeq$ *and* $\succcurlyeq^*$ *binary relations on* $\Delta(\mathbb{Y})\times\Delta(X)^N$ *and* $\Delta(X)$, *respectively, such that* $\succ^*\neq\emptyset$. *Then the following conditions are equivalent:*

(a) $\succcurlyeq$ *and* $\succcurlyeq^*$ *are complete, transitive, and continuous, jointly they satisfy consistency and* $\succcurlyeq$ *satisfies first-stage independence.*

(b) There exist continuous non-constant functions $V:\Delta(\mathbb{Y})\times\Delta(X)^N\to\mathbb{R}$ *and* $v:\Delta(X)\to\mathbb{R}$, *such that* V *represents* $\succcurlyeq$, v *represents* $\succcurlyeq^*$, *and, for all* $(\mu,\mathbf{p})\in\Delta(\mathbb{Y})\times\Delta(X)^N$,

$$V(\mu,\mathbf{p})=\sum_{y\in\mathbb{Y}}\mu(y)\,v(p(y)). \tag{8}$$

Moreover, if there are functions $\bar{V}:\Delta(\mathbb{Y})\times\Delta(X)^N\to\mathbb{R}$ *and* $\bar{v}:\Delta(X)\to\mathbb{R}$ *such that* $\bar{V}$ *represents* $\succcurlyeq$, $\bar{v}$ *represents* $\succcurlyeq^*$ *and* $\bar{V}(\mu,\mathbf{p})=\sum_{y\in\mathbb{Y}}\mu(y)\,\bar{v}(p(y))$ *then* $\bar{v}=bv+a,\ b>0$.

Proof: The proof is omitted, as it is identical to that of Theorem 1 of Karni and Safra (2000), with the additional assumption that the ex post preference relation $\succcurlyeq^*$ is signal-independent.

Remark: The behavior of a decision maker with such a pair of ex ante and ex post preference relations displays *dynamic consistency.* Formally, if $\mathbf{p}$ and $\mathbf{q}$ differ only when signal y occurs and $\mu(y)>0$, $p(y)$ is preferred to $q(y)$ ex post if and only if $\mathbf{p}$ is preferred to $\mathbf{q}$ ex ante.

The case in which the ex post preferences satisfy the independence axiom is of special interest. Let $U(p)=\Sigma_{x\in X}u(x)\,p(x)$ be an expected utility representation of $\succcurlyeq^*$. Then, by Theorem 1, the ex ante preferences $\succcurlyeq$ are represented by

$$V(\mu,\mathbf{p})=\sum_{y\in\mathbb{Y}}\mu(y)\,w(U(p(y))), \tag{9}$$

[9] This axiom is analogous to the constrained independence axiom defined in Karni and Safra (2000).

where w is strictly increasing. This representation is more general than expected utility (for $\succcurlyeq$) and is reduced to it when w is affine.[10] It resembles the class of non-expected utility models that have representations known as smooth ambiguity attitudes.[11] Note, however, that ambiguity aversion requires that the function w be concave. In the present context, a concave w implies aversion to the spread of signal-contingent payoffs that more informative experiments afford. See further discussion of this issue at the end of the next section.

5 Blackwell's Theorem Extended

Consider all pairs $(\succcurlyeq, \succcurlyeq^*)$ of ex ante and ex post preferences satisfying the requirements of Theorem 1(a). Let $\mathcal{V}$ be the set of all pairs (V, v) satisfying (8), and let $\mathcal{V}_{cx} \subset \mathcal{V}$ consist of all pairs in which v is convex (the need for convexity will be clarified in the next theorem). As above, for given $\pi \in \Delta(S)$, $B \subseteq 2^{\mathcal{H}} \backslash \varnothing$ and $\tilde{y} \in \mathcal{Y}$, let

$$f_y^* \in \arg\max_{f \in B} v\left(\Sigma_{s \in S} \pi(s \mid y) f(s)\right) \tag{10}$$

be the one-stage reduced lottery that maximizes the ex post utility v at the decision node associated with y. As the ex ante preferences $\succcurlyeq$ admit the hybrid representation and are dynamically consistent, the ex ante value of $\tilde{y}$ is given by $V(\mu, \left(f_y^*\right)_{y \in Y}) = \sum_{y \in Y} \mu(y) v(f_y^*)$. For every $((V, v), B) \in \mathcal{V}_{cx} \times 2^{\mathcal{H}} \backslash \varnothing$, the value of an experiment $\tilde{y} \in \mathcal{Y}$ with image Y is defined by:

$$\hat{V}(\tilde{y}; (V, v), \pi, B) := \sum_{y \in Y} \mu(y) v\left(f_y^*\right). \tag{11}$$

Definition 3: An experiment $\tilde{y}$ is *more hybrid-valuable* than another experiment $\tilde{y}'$ if, for all $((V, v), B) \in \mathcal{V}_{cx} \times 2^{\mathcal{H}} \backslash \varnothing$,

$$\hat{V}(\tilde{y}; (V, v), \pi, B) \geq \hat{V}\left(\tilde{y}'; (V, v), \pi, B\right).$$

The next theorem extends Blackwell's (1953) theorem.

Theorem 2: *An experiment $\tilde{y}$ is more hybrid-valuable than another experiment $\tilde{y}'$ if and only if $\tilde{y}$ is sufficient for $\tilde{y}'$.*

[10] In the social choice framework, Grant et. al (2010) provide an axiomatization of such preferences.
[11] See Klibanoff, Marinacci and Mukerji (2005) and Seo (2009).

The proof below rests on two properties of Blackwell's result: (i) that the expected utility under the sufficient experiment is larger and (ii) that the set of expected utility payoffs generated by the sufficient experiment constitutes a larger "spread" of expected utility values. With v monotonic increasing and convex, these properties make the expected value of v larger for the sufficient experiment.

Proof: (Sufficiency) Suppose that $\tilde{y}$ is sufficient for $\tilde{y}'$. Let Y and Y' denote the images of $\tilde{y}$ and $\tilde{y}'$, respectively. Fix $((V,v),B) \in \mathcal{V}_{cx} \times 2^{\mathcal{H}} \backslash \varnothing$. Then,

$$\begin{aligned}
\hat{V}\left(\tilde{y}';(V,v),\pi,B\right) &= \sum_{\{r|y'_r\in Y'\}} \mu'\left(y'_r\right) v\left(\sum_{s\in S} f^*_{y'_r}(s)\,\pi\left(s \mid y'_r\right)\right) \qquad (12)\\
&= \sum_{\{r|y'_r\in Y'\}} \mu'\left(y'_r\right) v\left(\sum_{s\in S} f^*_{y'_r}(s)\left[\sum_{\{j|y_j\in Y\}} \frac{\mu(y_j)\,m_{jr}}{\mu'(y'_r)}\pi(s \mid y_j)\right]\right)\\
&= \sum_{\{r|y'_r\in Y'\}} \mu'\left(y'_r\right) v\left(\sum_{\{j|y_j\in Y\}} \frac{\mu(y_j)\,m_{jr}}{\mu'(y'_r)}\left[\sum_{s\in S} f^*_{y'_r}(s)\,\pi(s \mid y_j)\right]\right)\\
&\leq \sum_{\{r|y'_r\in Y'\}} \mu'\left(y'_r\right) \sum_{\{j|y_j\in Y\}} \frac{\mu(y_j)\,m_{jr}}{\mu'(y'_r)} v\left(\sum_{s\in S} f^*_{y'_r}(s)\,\pi(s \mid y_j)\right)\\
&= \sum_{\{r|y'_r\in Y'\}} \sum_{\{j|y_j\in Y\}} \mu(y_j)\,m_{jr} v\left(\sum_{s\in S} f^*_{y'_r}(s)\,\pi(s \mid y_j)\right)\\
&\leq \sum_{\{r|y'_r\in Y'\}} \sum_{\{j|y_j\in Y\}} \mu(y_j)\,m_{jr} v\left(\sum_{s\in S} f^*_{y_j}(s)\,\pi(s \mid y_j)\right)\\
&= \sum_{\{j|y_j\in Y\}} \mu(y_j)\, v\left(\sum_{s\in S} f^*_{y_j}(s)\,\pi(s \mid y_j)\right) \sum_{\{r|y'_r\in Y'\}} m_{jr}\\
&= \hat{V}\left(\tilde{y};(V,v),\pi,B\right)
\end{aligned}$$

where the second equality follows from (5); the first inequality holds as v is convex; the second inequality holds since, by definition, $v(\Sigma_{s\in S} f^*_{y'_r}(s)\,\pi(s \mid y_j)) \leq v(\Sigma_{s\in S} f^*_{y_j}(s)\,\pi(s \mid y_j))$, and last equality holds as $\Sigma_{\{r|y_r\in Y'\}} m_{jr} = 1$.

(Necessity) If v is affine, then the expected utility representations are a subset of the set of preference relations that have hybrid representations with $(V,v) \in \mathcal{V}_{cx}$. Hence, necessity is implied by the necessity part of Blackwell's theorem. ♠

The ingredients of the representation of the hybrid model for affine v in (9) are similar to those of the class of non-expected utility models known as smooth ambiguity models.

Smooth ambiguity aversion requires that the analogue of the function w be concave. By contrast, our extension of Blackwell's theorem requires that decision makers exhibit *information proclivity*, an attitude captured by the convexity of the function w. This difference between the two models may be explained by the fact that more informative experiments offer larger spreads of expected utility payoffs to choose from. Consequently, in addition to the unambiguous advantage of allowing decision makers to choose acts whose payoffs are higher in the states that are more likely to obtain, information affects the decision-making process through the decision makers' attitudes toward wider spread of expected payoffs. We showed that the value of information increases if the decision maker displays information proclivity (that is, prefers a wider spread of expected payoffs). Information may be less valuable if the decision maker displays payoff-spread aversion, which is captured by concave w. The case of expected utility-maximizing behavior (i.e., the function w is linear) exhibits neutrality toward the payoffs spread.

Ambiguity aversion in smooth ambiguity models seems justifiable because decision makers are confronted with a spread of the priors that introduce payoff variations that they cannot exploit. By the same token, information proclivity is justifiable because decision makers value the opportunity to choose from, and thereby exploit, the wider spread of expected payoffs afforded by the more informative experiments.[12]

To grasp this point, let $S = \{s_1, s_2\}$, $Y = \{y_1, y_2\}$, $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$, and assume that v is strictly convex. Choose $p, q \in \Delta(X)$ such that $v\left(\frac{1}{2}p + \frac{1}{2}q\right) < \frac{1}{2}v(p) + \frac{1}{2}v(q)$, and consider $B = \{f\}$ where $f(s_1) = p$ and $f(s_2) = q$. Let $\tilde{y}$ be the fully informative experiment (that is, $\mu(y_j \mid s_j) = 1$, $j = 1, 2$) and $\tilde{y}'$ be the totally uninformative experiment ($\pi(\cdot \mid y_j) = \pi(\cdot)$, $j = 1, 2$). Clearly, $\tilde{y}$ is sufficient for $\tilde{y}'$. Then,

$$\hat{V}(\tilde{y}; (V, v), \pi, B) = \frac{1}{2}v(p) + \frac{1}{2}v(q) > v\left(\frac{1}{2}p + \frac{1}{2}q\right) = \hat{V}(\tilde{y}'; (V, v), \pi, B)$$

In addition to affording better choice of acts, information thus has intrinsic value in the sense that, ex ante, the decision maker prefers a wider spread of expected payoffs, even if their expectations is the same.

This example can be used to demonstrate the necessity of the convexity of v. Assume that v is not convex, and choose $p, q \in \Delta(X)$ such that $v\left(\frac{1}{2}p + \frac{1}{2}q\right) > \frac{1}{2}v(p) + \frac{1}{2}v(q)$. Then

$$\hat{V}(\tilde{y}; (V, v), \pi, B) = \frac{1}{2}v(p) + \frac{1}{2}v(q) < v\left(\frac{1}{2}p + \frac{1}{2}q\right) = \hat{V}(\tilde{y}'; (V, v), \pi, B)$$

and $\tilde{y}$, the fully informative experiment, becomes less hybrid-valuable than the totally uninformative experiment $\tilde{y}'$.

[12] Gensbittel, Renou and Tomala (2015) analyzed the amiguity case and provided conditions under which information is desirable within the maxmin model. Our main result does the same for the smooth ambiguity model. In both their paper and ours, the decision maker cannot commit ex ante to executing ex post choices.

Finally, the following example demonstrates the necessity of the first-stage independence axiom.

Example: Let $S = \{s_1, s_2\}$, $Y = \{y_1, y_2\}$ and $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$. Assume that v is a rank dependent utility functional defined by $v(p) = \int_{z \in X} z \mathrm{d}g(F_p(z))$, where F_p denotes the cumulative distribution functions of p and $g : [0,1] \to [0,1]$ is increasing and onto. For a two-outcome lottery $p = t\delta_x + (1-t)\delta_y$ $(x < y)$, $v(p) = g(t)x + (1-g(t))y$. It is known that v is convex if and only if g is concave.[13] Choose g to be the concave piecewise linear function

$$g(t) = \begin{cases} (2-\tau)t & t < 0.5 \\ 1 - \tau + \tau t & t \geqslant 0.5 \end{cases}$$

$(\tau < 1)$ and consider $B = \{f\}$ where $f(s_1) = \delta_1$ and $f(s_2) = \delta_2$. Let $\tilde{y}$ be the fully informative experiment and $\tilde{y}'$ be the less informative experiment satisfying $\mu'(y_1 \mid s_1) = 1$, $\mu'(y_2 \mid s_1) = 0$, $\mu'(y_1 \mid s_2) = \varepsilon$, $\mu'(y_2 \mid s_2) = 1-\varepsilon$, where $\varepsilon \in (0,1)$. Note that $\mu'(y_1) = \frac{1+\varepsilon}{2}$ and $\mu'(y_1) = \frac{1-\varepsilon}{2}$. Clearly, $\tilde{y}$ is sufficient for $\tilde{y}'$.

Assume that instead of using representation (8) of Theorem 1, we apply the rank-dependent utility v to the first-stage, too. Doing so is equivalent to using the certainty-equivalent reduction twice. Denoting the resulting value of experiments by $\check{V}$ and noting that $v(\delta_1) = 1$ and $v(\delta_2) = 2$, yields

$$\check{V}(\tilde{y}; (V, v), \pi, B) = g\left(\frac{1}{2}\right) v(\delta_1) + \left(1 - g\left(\frac{1}{2}\right)\right) v(\delta_2) = 2 - g\left(\frac{1}{2}\right) = 1 + \frac{\tau}{2}.$$

Similarly,

$$\begin{aligned} \check{V}\left(\tilde{y}'; (V, v), \pi, B\right) &= g\left(\frac{1+\varepsilon}{2}\right) v\left(\frac{1}{1+\varepsilon}\delta_1 + \frac{\varepsilon}{1+\varepsilon}\delta_2\right) + \left(1 - g\left(\frac{1+\varepsilon}{2}\right)\right) v(\delta_2) \\ &= g\left(\frac{1+\varepsilon}{2}\right)\left[g\left(\frac{1}{1+\varepsilon}\right) + \left(1 - g\left(\frac{1}{1+\varepsilon}\right)\right) 2\right] + \left(1 - g\left(\frac{1+\varepsilon}{2}\right)\right) 2 \\ &= 2 - g\left(\frac{1+\varepsilon}{2}\right) g\left(\frac{1}{1+\varepsilon}\right) = 2 - \left(1 - \tau + \tau\frac{1+\varepsilon}{2}\right)\left(1 - \tau + \tau\frac{1}{1+\varepsilon}\right). \end{aligned}$$

[13] Note that, by integration by parts, $v(p) = \int_{z \in X} z \mathrm{d}g(F_p(z)) = x_{\max} - \int_{z \in X} g(F_p(z)) \mathrm{d}z$. Hence, for a $\frac{1}{2} : \frac{1}{2}$ mixture and concave g,

$$\begin{aligned} v\left(\frac{1}{2}p + \frac{1}{2}q\right) &= x_{\max} - \int_z g\left(F_{\frac{1}{2}p + \frac{1}{2}q}(z)\right) \mathrm{d}z = x_{\max} - \int_z g\left(\frac{1}{2}F_p(z) + \frac{1}{2}F_q(z)\right) \mathrm{d}z \\ &\leqslant x_{\max} - \int_z \left[\frac{1}{2}g(F_p(z)) + \frac{1}{2}g(F_q(z))\right] \mathrm{d}z \\ &= \frac{1}{2}\left[x_{\max} - \int_z g(F_p(z)) \mathrm{d}z\right] + \frac{1}{2}\left[x_{\max} - \int_z g(F_q(z)) \mathrm{d}z\right] \\ &= \frac{1}{2}v(p) + \frac{1}{2}v(q) \end{aligned}$$

The converse is similarly proved.

For $\tau = 0.2$ and $\varepsilon = 0.1$, we get

$$\check{V}(\tilde{y}; (V, v), \pi, B) = 1.1 < 1.106 = \check{V}\left(\tilde{y}'; (V, v), \pi, B\right).$$

Hence, information may have negative value when first-stage independence is not assumed.

This example also demonstrates that employing the certainty-equivalent reduction twice does not eliminate aversion to information even if v is convex.

6 Concluding Remarks

6.1 The value of information

Lurking in the background of Blackwell's theorem are two tacit properties of expected utility theory: *consequentialism* and *reduction of compound lotteries*. The former maintains that, facing sequential decisions involving risky choices, decision makers are "forward looking" in the sense that, at every decision node, their preferences are unaffected by outcomes that did not materialize, ("roads not taken") along the decision-making path. The latter asserts that decision makers evaluate acts solely based on the ultimate probability distributions they induce on outcomes, regardless of whether the outcome is drawn, in a single step, from a known distribution by more convoluted trajectory that includes chance and decision nodes.

Wakker (1988) showed that departing from the independence axiom while maintaining consequentialism and reduction of compound lotteries (at all stages) and assuming quasi-convexity necessarily results in situations in which decision makers refuse free information related to finer partitions. Safra and Sulganik (1995) demonstrated that, maintaining consequentialism and reduction of compound lotteries throughout, violations of the relation of being more informative experiment à la Blackwell are quite robust. More specifically, for almost all pairs of experiments $\tilde{y}$ and $\tilde{y}'$ satisfying that $\tilde{y}$ is sufficient for $\tilde{y}'$, there exists a non-expected utility preference that strictly prefers the latter.

Replacing the RCLA by the certainty-equivalent reduction at all stages of the decision problem is not sufficient to derive an analogue of Blackwell's theorem (see Schlee, [1990] and Safra and Sulganik [1995]). This and the previous observations clarify the need for hybrid models.

6.2 Related literature

Segal (1990) was the first to propose a model of decision making under risk that replaces the reduction of compound lotteries with certainty equivalent reduction. Seo (2009) obtained smooth ambiguity averse representations of choice under uncertainty that departs from the reduction of compound lotteries axiom. Halevy (2007) presented experimental evidence suggesting that subjects whose behavior violates the reduction of compound lotteries under risk are more likely to exhibit ambiguity aversion when facing decision making under uncertainty.

This paper argues that departing from the independence axiom of expected utility theory while maintaining dynamic consistency and the intuitive presumption that being better informed is unambiguously desirable justifies the departure from the RCLA in sequential decision situations in which, at the interim stages, information may be exploited by choosing acts that better match the underlying data.

References

[1] Blackwell, D. (1953) "Equivalent Comparisons of Experiments," *Annals of Mathematics and Statistics*, 74, 265-272.

[2] Cremer, J. (1982) "A Simple Proof of Blackwell's "Comparison of Experiments" Theorem." *Journal of Economic Theory*, 27, 439-443.

[3] de Oliviera, Henrique (2018) "Blackwell's Informativeness Theorem Using Diagrams," *Games and Economic Behavior*, 109, 126-131.

[4] Gensbittel, F., Renou, L. and Tomala, T. (2015) "Comparisons of Ambiguous Experiments," Working Paper.

[5] Grant, S. Kajii, A. Polak, B. and Safra Z. (2010) "Generalized Utilitarianism and Harsanyi's Impartial Observer Theorem," *Econometrica,* 78(5):1939–1971.

[6] Halevy, Y. (2007) "Ellsberg Revisited: An Experimental Study," *Econometrica*, 75, 503–536.

[7] Karni, E. and Safra, Z. (1989) "Dynamic Consistency, Revelation in Auctions and the Structure of Preferences." *Review of Economic Studies* 56, 421-433.

[8] Karni, E. and Safra, Z. (2000) "An Extension of a Theorem of von Neumann and Morgenstern with an Application to Social Choice Theory," *Journal of Mathematical Economics*, 34, 315-327.

[9] Karni, E. and Schmeidler D. (1991) "Atemporal Dynamic Consistency and Expected Utility Theory." *Journal of Economic Theory* 54, 401–408.

[10] Klibanoff, P., Marinacci, M. and Mukerji, S. (2005) "A Smooth Model of Decision Making Under Ambiguity," *Econometrica*, 73 , 1849-1892

[11] Leshno, M. and Spector, Y. (1992) "An Elementary Proof of Blackwell' Theorem." *Mathematical Social Sciences*, 25, 95-98.

[12] Li, J. and Zhou, J. (2016) "Blackwell's Informativeness Ranking with Uncertainty Averse Preferences," *Games and Economic Behavior,* 96, 18-29.

[13] Safra, Z. and Sulganik, E. (1995) "On the Nonexistence of Blackwell's Theorem-Type Results with General Preference Relations." *Journal of Risk and Uncertainty* 10, 187-201.

[14] Schlee, E. (1990) "The Value of Information in Anticipated Utility Theory," *Journal of Risk and Uncertainty,* 3, 83-92.

[15] Segal, U. (1990) "Two Stage Lotteries without the Reduction Axiom." *Econometrica*, 58, 349–377.

[16] Seo, K. (2009) "Ambiguity and Second-Order Beliefs," *Econometrica*, 77, 1575–1605.

[17] Wakker, P. (1988) "Nonexpected Utility as Aversion to Information," *Journal of Behavioral Decision Making* 1, 169-175.