Moral Hazard, the Savage Framework, and State-Dependent Utility*

Jean Baccelli†

Munich Center for Mathematical Philosophy

February 7, 2019

Abstract

In this paper, I investigate the betting behavior of a decision-maker who can influence the likelihood of the events upon which she is betting. In decision theory, this is best known as a situation of moral hazard. Focusing on a particularly simple case, I sketch the first systematic analysis of moral hazard in the canonical Savage framework. From the results of this analysis, I draw two philosophical conclusions. First, from an observational and a descriptive point of view, there need to be no incompatibility between moral hazard and the Savage framework. This qualifies the incompatibility view, that is ubiquitous in decision theory. Second, in general, moral hazard is not sufficient to overcome the challenges posed by state-dependent utility to the behavioral identification of beliefs. This qualifies the sufficiency view, that is influential in decision theory. These two philosophical conclusions are the main contributions of my paper.

*For helpful comments, I am grateful to two reviewers, Richard Bradley, Eric Danan, Raphaël Giraud, Edi Karni, Yang Liu, Philippe Mongin, Jiji Zhang, and Fanyin Zheng. I also thank audiences at the Department of Economics at the Université de Cergy-Pontoise, the Departments of Philosophy at the Università degli Studi di Milano and the Universität Bayreuth, and the Department of Statistics at the Ludwig-Maximilians-Universität München. All errors and omissions are mine. Research for this paper was funded by the Wagemann Foundation and the Ludwig-Maximilians-Universität München.

†jean (dot) baccelli [at] gmail (dot) com

Introduction

Consider a decision-maker whose actions can affect the likelihood of some events. For instance, consider a manager and the event that one of her employees be promoted. Given her direct authority over the matter, the manager might have outright *control* over this event, i.e., be able to make it occur with any probability ranging from 0 to 1. Similarly, she might have at least some *influence* over the event that the employee of one of her colleagues' be promoted, i.e., be able to make it occur with any probability in some strict sub-interval of $[0, 1]$. For instance, it might be by default equally likely than not that the promotion will occur. But she might be able to provide her colleague with persuasive recommendations in one direction or the other, with the result of making the event occur with, say, any probability ranging from $\frac{1}{4}$ to $\frac{3}{4}$. Some other promotions might, by contrast, be entirely out of the reach of her influence, e.g., because she does not know the relevant manager. These definitions for "control" and "influence" generalize from one event to a set of events. Consequently, a decision-maker can be understood as having full, partial, or no control (respectively, influence) over an algebra of events, thus covering a rich variety of concrete situations.[1]

Consider a decision-maker having at least partial influence over an algebra of events—full control thus being a limit case, rather than the exclusive case of interest. Assume that an outside observer has as a goal to identify the beliefs which the decision-maker holds, e.g., under the form of subjective probability values, regarding the likelihood of these events. The observer might have this goal for several reasons—for instance, he wants to use these beliefs in his own decision-making. Crucially, assume that he cannot observe the actions by which the decision-maker can exert her influence over the events. In terms of the example above, this is the case if, e.g., he is simply unaware of the fact that the decision-maker is the manager of the first employee and able to give recommendations to the manager of the second. Assume that all he can observe is the *betting behavior* of the decision-maker. By betting behavior, I mean her choices between all conceivable bets on the events in question. I understand bets abstractly, as functions from the set of events (here, the various promotions possible) into some set of payoffs between which she proves not indifferent (e.g., various sums of money).[2]

As the following example suggests, the payoffs of the bets will give the decision-maker incentives to exert her influence in different directions. If

[1]This contrast between control and influence is taken from Drèze and Rustichini, 1999. The convexity assumption (all values in the interval) can be considered an idealization.

[2]Given the topic under consideration, it is best to depart from traditional terminology and speak of "bets" instead of (Savagian) "acts". Otherwise, in general, one would need to distinguish between "acts" and "actions", which is unnatural and cumbersome (more on this in fn. 8). Relatedly, in my paper, "bets" will be understood as generally as (Savagian) "acts" usually are, not as restrictively as some decision theorists recommend. I have in mind those for whom bets are *binary* acts (see, e.g., Ghirardato and Marinacci, 2001, p. 865).

presented with a bet offering her \$100 if her employee is promoted, \$0 otherwise, the decision-maker will, ceteris paribus, have an incentive to promote her employee so as to collect the \$100. Inversely, if the payoffs of this bet are swapped, she will have an incentive not to promote her employee. This will, in various ways to be detailed later, complicate the identification exercise. The complication is reminiscent of the so-called *moral hazard* cases investigated in several branches of economics, most famously in contract theory.[3] Here is a typical example, coming from the theory of insurance contracts. By offering excessively high indemnities in the event of a house fire, an insurance company might induce its insurees into behaving more carelessly than otherwise, sometimes, even into setting their own houses on fire in order to collect the money. This would affect the probability of a house fire upon which the company had first calculated its insurance offer, therefore, its profits. Admittedly, although an insurance policy is comparable to a bet, there is an important difference between the typical moral hazard cases, that belong to game theory, and the situation under consideration here, that belongs to decision theory. Unlike an insurance company contracting with an insuree, an outside observer trying to analyze the decision-maker's betting behavior has no real stake in the situation, such as a profit function. By assumption, his only goal is to correctly identify her beliefs. This asymmetry notwithstanding, it is best to follow established terminology in categorizing the decision-theoretic situation under consideration as one of moral hazard.

Accordingly, within the present paper, call *problem of moral hazard* the problem consisting in identifying the beliefs which a decision-maker holds about some uncertain events, merely by observing how she would bet on these events, when, by some unobservable side actions, she can also influence the resolution of uncertainty, therefore, the likelihood of the events.[4] My paper is concerned with investigating the problem of moral hazard, both mathematically and conceptually, in the canonical decision-theoretic framework of Savage (Savage, 1954, with a second revised edition in 1972). As customary in decision theory, I distinguish between Savage's framework and Savage's own decision model within it, i.e., the classical subjective expected utility model. At this stage, suffice to say that the Savage framework exactly corresponds to the abstract betting environment just invoked. In this framework, only the betting behavior of the decision-maker is assumed observable. Besides, the framework is "purely subjective", in the sense that it neither explicitly nor implicitly lists any probabilistic notion among its observational primitives.

[3]See, e.g., Hart and Hölmstrom, 1987. The kind of moral hazard problem that is closest to the decision theory problem examined in the present paper is the classic "hidden action" problem (see, e.g., Hart and Hölmstrom, 1987, p. 76).

[4]More generally, the problem arises if the decision-maker *thinks* that her actions can influence the likelihood of the events, i.e., even if such is, in fact, not the case. Similarly, the standard no-influence case corresponds, in effect, to when the decision-maker thinks, rightly or wrongly, that none of her actions can influence the likelihood of the events.

Surprisingly, the decision-theoretic literature contains no systematic investigation of the problem of moral hazard in the canonical Savage framework. The two main pre-existing systematic investigations of this problem, due to Drèze (see especially Drèze, 1987a; also Drèze, 1961, as translated in Drèze, 1987b; Drèze and Rustichini, 1999, Drèze and Rustichini, 2004) and Karni (see especially Karni, 2011a, Karni, 2011b; also Karni, 2006, Karni, 2013), are both set in a non-Savagian framework.

Drèze's analysis is conducted in the so-called "Anscombe-Aumann framework" (Anscombe and Aumann, 1963). In this framework, bets are defined not directly into the set of payoffs, but into the set of all possible probability distributions over the set of payoffs. There is not only, in general, an unwelcome "conceptual dissonance" (Alon and Schmeidler, 2014, p. 384) in thus trying to analyze subjective probabilities by taking some probability values for granted. There are also reasons, specific to the context of moral hazard, for considering this approach as too restrictive. The probability values that are taken for granted introduce a sub-class of events that, by definition, cannot be under the influence (let alone the control) of the decision-maker.

On the other hand, unlike Drèze, Karni does not take any probability value for granted. But the original framework in which his analysis is conducted suffers from another limitation. Unlike Drèze, Karni supposes that the actions by which the decision-maker can influence the likelihood of the events are observable. As Karni's results illustrate, this significant enrichment of the Savage framework does not, in the least, trivialize the identification exercise. However, it simply does not square with the problem of moral hazard as I have stated it. Besides, although Karni's approach is in some respects the closest to the economics of moral hazard, it is at variance with it in this particular respect. In the economics of moral hazard, the relevant actions are supposed impossible—alternatively: too costly—to observe.[5]

I have two main motivations for following in the footsteps of Drèze and Karni by investigating the problem of moral hazard in the Savage framework. My main contributions can be stated with reference to each of them.

First, I want to assess the widespread claim according to which there is an incompatibility between moral hazard and not just Savage's own decision model, but his very framework. This claim is found most easily in the

[5]Karni's approach is nonetheless the closest to the economics of moral hazard inasmuch as (as I discuss in more detail on p. 6), unlike Drèze's approach, it accommodates actions with variable intrinsic utilities. Arguably, this is the main source of the incentive compatibility problem central to the economics of moral hazard. Regarding the unobservability issue, Karni writes (Karni, 2006, p. 335): "[t]he position taken here is that, whether or not actions are observable by a second party, decision-makers are aware of the actions they may take and have well-defined preference relations on action-bet pairs". But thus retreating to the *introspective* point of view is in tension with emphasizing, which Karni justifiedly does, the significance of his identification results from the *observational* point of view. It would be particularly interesting to try to generalize Karni's approach by supposing that the actions are observable, but only imperfectly so (as, e.g., choice probabilities could express).

parts of decision theory that are the closest to philosophy, especially in the so-called "causal decision theory" literature. For instance, at the beginning of an influential contribution to this literature, Joyce writes (Joyce, 1999, p. 57; emphasis in original): "Savage's framework (...) makes sense only when the probabilities of the states (...) are *independent* of the decision-maker's (...) choice of actions." But the claim can also be found in the parts of decision theory that are the closest to economics. For instance, at the beginning of his authoritative textbook on so-called "prospect theory", Wakker writes (Wakker, 2010, p. 32-33): "we require that the decision-maker does not have any influence on the truth of the events. (...) The techniques of this book[, which is essentially set in the Savage framework,] cannot be applied directly to such cases." Typically, when endorsing this incompatibility claim, decision theorists—including Wakker, in the specific passage quoted above—consider decision-making from an introspective and a prescriptive perspective, rather than from the observational and the descriptive perspective built in my statement of the problem of moral hazard. Their core claim seems to be that, under moral hazard, the Savage framework is inadequate for capturing our considered intuitions about which decisions are normatively compelling. However, the exact scope of the claim is left open. In particular, the core claim above is not clearly distinguished from the claim according to which, under moral hazard, the Savage framework is inadequate for articulating any observationally well-defined model of decision-making. On the contrary, I take the above passages to suggest that, whenever the resolution of uncertainty is not entirely exogenous to the decision-maker's choices, the Savage framework—by contrast with, say, Jeffrey's framework (Jeffrey, 1965, with a second revised edition in 1983)—is ill-suited for doing any sort of decision theory under uncertainty.

My contribution will be to show that, from an observational and a descriptive point of view, there need to be no incompatibility between moral hazard and the Savage framework. To show this, I will build on the fact that a particularly simple form of moral hazard proves dual to the familiar "max-min expected utility" model of Gilboa and Schmeidler (Gilboa and Schmeidler, 1989). The duality with such a well-behaved decision model importantly qualifies the incompatibility view, by clarifying that it must have a restricted scope. Indirectly, this also contributes to a better understanding of the nature of the issues examined in the causal decision theory literature.

Second, I want to assess a remarkable claim coming from the literature dedicated to the problem of state-dependent utility. This other problem is as general as, but distinct from, the problem of moral hazard.[6] In a nutshell (details will follow), the problem stems from the fact that the preferences of the decision-maker can vary with the events on which she is betting, in

[6]Because of their comparable generality, they are often presented as "twin" problems for decision theory under uncertainty (see, e.g., Drèze, 1987a, p. 75).

ways unrelated to, but nonetheless impeding the identification of, her beliefs about the respective likelihoods of these events. It is now a well-established conclusion in the literature that the possible state-dependence of utility is the source of major complications for the behavioral identification of beliefs (see, e.g., Schervish et al., 1990; Karni, 1996). However, both Drèze and Karni, who are also leading experts on the identification issues associated with state-dependent utility, claim that for these issues to be solved, it suffices to suppose that the decision-maker has the capacity to influence the likelihood of the events, as in situations of moral hazard (see, e.g., Drèze, 1987a, p. 29; Karni, 2008, p. 228 sq.). Such situations would even be, as it is claimed, the only ones in which beliefs can be given satisfactory behavioral—a.k.a. revealed preference—foundations, i.e., fully identified with no introspective input whatsoever. In light of the importance of such revealed preference methodology in decision theory, this claim is methodologically remarkable, and well worth assessing in the canonical Savage framework.

My contribution will be to show that, in general, moral hazard does not suffice to eschew the identification issues raised by state-dependent utility. I will offer two arguments to establish this non-sufficiency. One stems from technical considerations, which Drèze has anticipated, but—or so I will argue—misappreciated in one crucial respect. The other argument stems from philosophical considerations, which, I will explain, neither Drèze nor Karni have anticipated, let alone addressed. Establishing the above non-sufficiency claim contributes to the philosophical assessment of the revealed preference methodology and, thereby, the philosophy of economics at large.

The paper is organized as follows. The first section sketches the axiomatic analysis of the simple form of moral hazard fitting my purposes. This is where I notice and exploit the duality with the max-min expected utility model, thus establishing that some forms of moral hazard are fully expressible in the Savage framework. The second section provides a more conceptual discussion of the decision model thus obtained. This is where I show that for several reasons, this model does not satisfactorily identify the underlying beliefs of the decision-maker, thus establishing that some forms of moral hazard leave the problem of state-dependent utility unsolved. A brief conclusion ensues.

1 Axiomatic analysis of moral hazard

Let S be the set of all possible states of nature, on which the decision-maker is to place her bets. States are exclusive of one another, and one and only one of them, denote it by s^*, is the true state of nature. Before the bets are placed, it is not known which $s \in S$ is s^*; but this can be objectively verified afterwards. Let Σ stand for the σ-algebra of all possible events, i.e., sets of possibles states of nature. Let X be the set of all possible payoffs of the bets. For concreteness, also because of the richness assumptions made in the specific result which I will build on, assume that X is, say, a monetary interval, for instance, $[0, 100]$.[7] Let F stand for the set of all possible bets, i.e., functions $f : S \to X$ that are measurable with respect to Σ and take finitely many values over S. The bets taken by the decision-maker are assumed observable and representable by a binary relation $\succcurlyeq$ over F, with $\sim$ and $\succ$ its symmetric and asymmetric part, respectively. There is also a set A of actions by which the decision-maker thinks that she can influence the likelihood of at least some events. But, unlike bets, actions are not assumed observable. Accordingly, their role in subsequent analysis will, typically, be left implicit.[8]

As I have started suggesting in the introduction (when highlighting the distinction between control and influence), moral hazard can come in various forms. The specific form investigated in this paper is the simplest of all, inasmuch as it is characterized by the following assumptions.

First, once and for all in the present section, assume away any form of state-dependent utility. Formally, this means that in the representation to be constructed, the betting decisions will be driven by a utility function u, the domain of which is X, not $X \times S$. In order to study the interaction between the problem of moral hazard and the problem of state-dependent utility, which I will do in the next section, it is best to first keep them separate.

Second, assume away any variable intrinsic utility or disutility of the actions by which the decision-maker (thinks that she) can exert her influence over the events. Essentially, this means that in the representation, the domain of the utility function u will indeed be X, not $X \times A$. This does not exclude that the set A corresponds to different levels of effort on behalf of the decision-maker. But this does exclude that these different levels of effort are, in one way or another, associated with different utility levels. As the economic literature on moral hazard illustrates, this is a major assumption. While it would not be acceptable for many purposes, it is so for the conceptual purposes of this paper, viz., establishing that at least one form of moral hazard is compatible with the Savage framework and does not eschew the identification issues raised by state-dependent utility. The assumption that actions do not have variable intrinsic (dis)utilities also governs Drèze's analysis, while such is not the case of Karni's analysis. This may be the most

[7] More generally, I will assume (as in A0, to be introduced later) that X is a connected topological space. By contrast and as a result, unlike in most other investigations of the Savage framework, I will not need to make any richness assumption regarding S.

[8] Given the topic under investigation, it is best to depart from standard terminology, which consists in calling F the set of "acts" and X, the set of "consequences". In general, this tends to unduly constrain the admissible interpretations of Savage's framework. In addition, in the particular case of moral hazard, this also crushes some potentially relevant distinctions. Crucially, it is not *by betting* on some event that one makes it more or less likely to occur, but by some other action that is entirely distinct from the betting itself. Admittedly, one could stick to standard terminology and distinguish between "actions" and "acts". But this distinction strikes me as unnatural, and cumbersome to keep track of.

significant difference between these two analysis. It can be related to the fact that Karni, unlike Drèze and myself, supposes the actions observable. It is in general difficult, and conceivably impossible in some instances, to infer from some observations (here, a series of bets) both an unobservable set (a set of actions parallel to the bets) and a form of measurement over that set (a utility function over this set of actions). Be it only for this reason, it is wise to start investigating moral hazard in Savage's framework by first excluding any variable intrinsic (dis)utility of the unobservable actions.[9]

Third, assume the simplest departure from Bayesianism applicable to our case, which covers the following sub-assumptions. First, each action induces a subjective probability measure π over the set of events, thus generating a set Π of subjective probability measures. Second, while only her choice between the bets is observable, what the decision-maker truly does is as follows. Given any bet, she chooses the action that will induce the maximal expected utility associated with this bet. Given any pair of bets, she chooses the bet to which, together with some action, the maximal expected utility is associated. Formally, for some suitably unique utility function $u : X \to \mathbb{R}$ and set Π of probability measures $\pi : \Sigma \to [0, 1]$, and with $\mathbb{E}^u_\pi(f)$ denoting the expectation associated with f of u with respect to π, our target is a representation of $\succcurlyeq$ by some function $v : F \to \mathbb{R}$ that can be analyzed as follows:

$$v(f) = \max_{\pi \in \Pi} \mathbb{E}^u_\pi(f). \tag{1}$$

Under an entirely different interpretation (and some further restrictions, e.g., the convexity of Π), this representation happens to be well known, which I will exploit shortly. But first, it is enlightening to check the implications of moral hazard, understood as in (1), regarding the most important axioms of the Savage theorem, i.e., the axiomatization of the traditional subjective expected utility model, or (1) when Π is a singleton. This is enlightening because, however clear they might be mathematically, as when one starts directly from (1), these implications are non-trivial conceptually, if one starts, instead, from the bare idea that the decision-maker can exert some unobservable influence over the events upon which she is observed to bet. This will also prepare the discussion of state-dependent utility in the next section.

The axioms of the Savage theorem which I want to focus on pertain to the construction, based on the observation of the bets taken by the decision-maker, of a so-called "qualitative probability relation".[10] More precisely, they consist in postulating that the bets taken by the decision-maker induce a

[9] Notice that in the claim according to which moral hazard is incompatible with the Savage framework, the fact that actions may have variable intrinsic utilities plays no role.

[10] For the exact definition of a qualitative probability relation, see, e.g., Fishburn, 1970, p. 195. More generally, see the whole Ch. 14 for the standard proof of the Savage theorem.

well-defined likelihood order over the set of events (P4), and that this order has probabilistic properties, inasmuch as conditional betting is well-defined, i.e., independent of the payoffs outside of the conditioning event (P2), and, when restricted to constants, essentially independent of the conditioning event (P3). Formally, with $f, g \in F$, let fEg stand for the bet, the payoff of which is identical to that of f if the true state of nature is in E, that of g otherwise. With $x \in X$, whenever applicable, let x stand for the constant bet, the payoff of which is x, whatever the true state of nature (thus, x can be understood as the bet xEx, for any $E \in \Sigma$). Define an event E as non-null if there exist $x, y \in X$, and $h \in F$, such that $xEh \succ yEh$, and define it as null otherwise. Then, the relevant axioms of the Savage theorem read as follows.

P4 $\quad \forall E, E' \in \Sigma, \forall x, y, x', y' \in X$ such that $x \succ y$ and $x' \succ y'$:
$xEy \succcurlyeq xE'y \Leftrightarrow x'Ey' \succcurlyeq x'E'y'$.

P2 $\quad \forall E \in \Sigma, \forall f, g, h, h' \in F : fEh \succcurlyeq gEh \Leftrightarrow fEh' \succcurlyeq gEh'$.

P3 $\quad \forall$ non-null $E \in \Sigma, \forall x, y \in X, \forall h \in F : x \succcurlyeq y \Leftrightarrow xEh \succcurlyeq yEh$.

Consider first the implications of our simple form of moral hazard as regards P4. One can prove that (1) respects P4. This is shown by the observation that (1) induces a well-defined likelihood order over Σ, namely, the order representable by the function $\lambda : \Sigma \to [0, 1]$ defined as follows: $\forall E \in \Sigma$, $\lambda(E) = \max_{\pi \in \Pi} \pi(E)$.[11] The fact that P4 is respected here might be counter-intuitive to some. This is because P4 is often interpreted as imposing that the probability of an event be independent of the payoff associated with it in any particular bet offered to the decision-maker (see, e.g., Ellsberg, 1961, p. 649), and moral hazard is often glossed as a form of payoff-dependent probability (see, e.g., Drèze, 1961, p. 77, as translated in Drèze, 1987b, p. 94). However, one can see from the above that P4 is respected by at least one form of moral hazard, namely, the simple form in (1).[12]

Consider next the implications of this form of moral hazard as regards P3. One can prove that (1) need not, but can, violate P3, and, if at all, in a very distinctive way (on what follows, see, e.g., Drèze, 1987a, p. 55). To see why,

[11]See Epstein and Le Breton, 1993, p. 13 for a similar observation. It can be proved as follows. By (1), $xEy \succcurlyeq xE'y$ if and only if $\max_{\pi \in \Pi} \mathbb{E}^u_\pi(xEy) \geq \max_{\pi \in \Pi} \mathbb{E}^u_\pi(xE'y)$. Developing, this holds if and only if $\max_{\pi \in \Pi} \left[\pi(E)\big(u(x) - u(y)\big) + u(y)\right] \geq \max_{\pi \in \Pi} \left[\pi(E')\big(u(x) - u(y)\big) + u(y)\right]$. Because it is assumed that $u(x) - u(y) > 0$, this simplifies to $\max_{\pi \in \Pi} \pi(E) \geq \max_{\pi \in \Pi} \pi(E')$. The claim is thus proved, because the inequality does not depend on the particular x and y such that $u(x) - u(y) > 0$. Importantly, notice that $\max_{\pi \in \Pi} \pi(E)$ and $\max_{\pi \in \Pi} \pi(E')$ may be reached by different elements $\pi \in \Pi$.

[12]Some other forms of moral hazard violate P4—see Appendix A.1 for a simple example.

recall the example from the introduction. For the present discussion and future reference, let E_1 be the event that the employee of the decision-maker is promoted, E_2, the event that the employee of her colleague is promoted, E_3, the event that the employee of some unknown manager is promoted, and assume that all these events are mutually compatible. As in the introduction, assume that the decision-maker can make E_1 occur with any probability in $[0, 1]$ (a case of control, as defined in this paper), E_2, with any probability in $[\frac{1}{4}, \frac{3}{4}]$ (a case of influence), and that she cannot affect the probability of E_3, which she believes to be fixed at $\frac{1}{2}$. Then, making merely the ordinal assumption that she has an increasing utility function for money, she will make the following betting decisions, which taken together constitute a violation of P3: $100E_1 50 \succ 0E_1 50$, $50 \succ 0$ but $50E_1 100 \sim 0E_1 100$. The first choice indicates that the decision-maker treats E_1 as non-null. The second expresses that she generally prefers more money to less. The third is explained by the fact that given either bet in the particular pair she is presented with (i.e., $50E_1 100$ and $0E_1 100$), it is optimal for her to push the probability of E_1 to 0—which, having control over E_1, she can do—and collect the \$100.

Noteworthily, control is not necessary for violations of P3 such as the one just introduced to occur. It suffices that there be one event over which the decision-maker has some influence reaching probability 0, i.e., which she can make occur with either probability 0, or probability a, with $0 < a \leq 1$ (and, possibly, with all the intermediary values in $(0, a)$). The choice pattern thus isolated is the only form under which P3 can be violated by (1).[13] This is because the specificity of (1) is, in a nutshell, entirely on the probability side. Probability values being positive numbers (so that two probability values can always be related by some increasing transformation), no strict preference reversal can occur, as when for some $x, y \in X$, $E \in \Sigma$, and $h \in F$, $x \succ y$ is observed together with $xEh \prec yEh$.

These observations regarding P3 and P4 lead to a valuable intermediary conclusion. As is well known and I will emphasize again in the next section, P3 and P4 are the state-independent utility axioms of the Savage axiomatics. If the (potentially) state-dependent utilities underlying the choices of the decision-maker could be elicited separately from her beliefs, with u_E and $u_{E'}$ the resulting functions corresponding to events E and E', P3—respectively, P4—would consist in requiring that any pair $u_E, u_{E'}$ be related by an increasing—respectively, affine—transformation. Essentially, one would thus require that u_E and $u_{E'}$ express the same basic ordering (P3) over the set of payoffs, as well as the same risk attitudes (P4) with respect to that ordering. Now, P4 is respected by (1). P3 can be violated by (1). But the violations take a distinctive form that cannot be interpreted as an instance of state-dependent utility. This is because, as next paragraph will confirm,

[13]It can be restated by noting that decision models such as (1) respect only some forms of state-wise or stochastic dominance (more details in, e.g., Wakker, 2010, App. 11.10).

the key feature of these cases is not what happens *on* the conditioning event, by contrast with some other conditioning event, but what happens *outside* the conditioning event, i.e., on its complementary event. All this leads to the non-trivial conclusion according to which, under the assumptions listed at the beginning of this section, *state-dependent utility and moral hazard are behaviorally disjoint phenomena.*[14]

Finally, consider the implications of the form of moral hazard in (1) as regards P2. One can prove that (1) typically violates P2 (on what follows, see, e.g., Drèze and Rustichini, 2004, p. 865). An illustrative violation is displayed next, still in terms of the example formally introduced on p. 8, and with the customary symbols for intersection and complementation:

$$\underbrace{\begin{bmatrix} 50 & \text{if } s^* \in E_3 \cap E_1 \\ 0 & \text{if } s^* \in E_3 \cap \overline{E_1} \\ 40 & \text{if } s^* \in \overline{E_3} \cap E_1 \\ 30 & \text{if } s^* \in \overline{E_3} \cap \overline{E_1} \end{bmatrix}}_{f} \succ \underbrace{\begin{bmatrix} 0 & \text{if } s^* \in E_3 \cap E_1 \\ 50 & \text{if } s^* \in E_3 \cap \overline{E_1} \\ 40 & \text{if } s^* \in \overline{E_3} \cap E_1 \\ 30 & \text{if } s^* \in \overline{E_3} \cap \overline{E_1} \end{bmatrix}}_{g} \;\&\; \underbrace{\begin{bmatrix} 50 & \text{if } s^* \in E_3 \cap E_1 \\ 0 & \text{if } s^* \in E_3 \cap \overline{E_1} \\ 10 & \text{if } s^* \in \overline{E_3} \cap E_1 \\ 100 & \text{if } s^* \in \overline{E_3} \cap \overline{E_1} \end{bmatrix}}_{f'} \prec \underbrace{\begin{bmatrix} 0 & \text{if } s^* \in E_3 \cap E_1 \\ 50 & \text{if } s^* \in E_3 \cap \overline{E_1} \\ 10 & \text{if } s^* \in \overline{E_3} \cap E_1 \\ 100 & \text{if } s^* \in \overline{E_3} \cap \overline{E_1} \end{bmatrix}}_{g'}.$$

Example 1 – A violation of P2 due to moral hazard

Focus on f in Example 1. Recall that E_1 is under the control of the decision-maker, while E_3 is not even under her influence. Facing f, the only relevant question for the decision-maker is what is the optimal value $\pi(E_1)$ for E_1, and it suffices here that she focuses on values 0 and 1. If she pushes $\pi(E_1)$ to 0, i.e., does not promote her employee, she in effect turns f into an equiprobable gamble with payoffs \$0 and \$30. If, instead, she pushes $\pi(E_1)$ to 1, i.e., promotes her employee, she in effect turns f into an equiprobable gamble with payoffs \$50 and \$40. Obviously, under the assumption that she has an increasing utility function for money, she should push $\pi(E_1)$ to 1. Invoking such first-order stochastic dominance considerations (when necessary, across bets) with reference to (1), it follows that $f \succ g$ and $f' \prec g'$, a violation of P2—witness the pairwise common payoffs on $\overline{E_3}$.

Importantly, it is inessential to the example that it involves an event, E_1, over which the decision-maker has control. For instance, applying the same kind of first-order stochastic dominance reasoning, one can check that the violation of P2 in Example 1 still arises with E_1 being replaced by E_2, i.e., an event over which the decision-maker has some influence, but not control.

[14] The conclusion is non-trivial be it only because, as illustrated in Appendix A.1, it does not hold in general. Specifically, under some forms of moral hazard, P4 can be violated because of either moral hazard, or state-dependent utility, or both.

It should not come as a surprise that, when understood as in (1), moral hazard violates P2. (1) is dual to one of the best-known models of decision theory, namely, the so-called "max-min expected utility" model of Gilboa and Schmeidler (Gilboa and Schmeidler, 1989). This model has been specifically designed to accommodate violations of P2, such as the Ellsberg paradox (Ellsberg, 1961). In the max-min expected utility model, with the same kind of underlying parameters and notation as in (1), the betting decisions of the decision-maker under consideration are represented by the following function:

$$w(f) = \min_{\pi \in \Pi} \mathbb{E}^u_\pi(f). \tag{2}$$

(1) is dual to (2) inasmuch as max-max is dual to max-min. Surprisingly, this duality between max-min expected utility and one form of moral hazard has been rarely noted and, more importantly, never elaborated upon or exploited.[15]

This duality can be exploited for my purposes. While axiomatizations of max-min expected utility have long been restricted to the Anscombe-Aumann framework or comparable settings, Alon and Schmeidler have recently proposed an axiomatization in Savage's purely subjective environment (Alon and Schmeidler, 2014).[16] I now indicate and discuss the modification that is necessary and sufficient for Alon and Schmeidler's theorem to yield an axiomatization of (1).

The content of the relevant modification is explained by the nature of the difference between (1) and (2). The difference does not reside in the existence of parameters u and Π, i.e., the so-called "multi-prior" structure. It resides in how these parameters are combined in evaluating the bets. While (2) embodies a worst-case scenario evaluation, (1) embodies a best-case scenario evaluation, as befits the decision-maker's alleged capacity to make the best-case scenario happen. Accordingly, for my purposes, the relevant modification is as follows. Alon and Schmeidler's "uncertainty aversion" postulate needs to be revised into its dual, i.e., an "uncertainty seeking" postulate.

To state the revised postulate in terms of the primitive $\succcurlyeq$, I need some additional definitions and notation. Recall that two bets f and g are called co-monotonic for a given decision-maker if and only if there is no $s, t \in S$ such that $f(s) \succ f(t)$ and $g(s) \prec f(t)$. In other words, f and g induce the same ordering of the states, as induced by the ordering of the payoffs which f and g associate to each state. From the primitive binary relation $\succcurlyeq$ over F, derive a quaternary relation $\widehat{\succcurlyeq}$ over X, with symmetric and asymmetric part $\widehat{\sim}$ and $\widehat{\succ}$, based on the following definition: $wx \widehat{\succcurlyeq} yz$ if and only if for some $a, b \in X$ and $E \in \Sigma$, $wEa \succcurlyeq xEb$ and $yEa \preccurlyeq zEb$, with the four bets involved co-monotonic, and E non-null on the largest set of co-monotonic

[16] More specifically, the axiomatization concerns (2) with Π a closed, convex set of finitely additive probability measures on Σ, and u a continuous utility function on X.

bets including these four bets. The other postulates of Alon and Schmeidler, recalled in Appendix A.2, ensure that this derived quaternary relation $\widehat{\succcurlyeq}$ is well-behaved. They also ensure that, when defined over a rich enough X, it delivers the utility parameter u featured in (1), with $wx\widehat{\succcurlyeq}yz$ holding if and only if $u(w) - u(x) \geq u(y) - u(z)$ also holds.[17] It is useful to remark that, in particular, $xy\widehat{\sim}yz$ holds if and only if $u(y) = \frac{1}{2}u(x) + \frac{1}{2}u(z)$.

The signature axiom of moral hazard can now be introduced:

AX $\quad \forall f, g, h \in F:$ if $f \succcurlyeq g$, and $f(s)h(s)\widehat{\sim}h(s)g(s)\ \forall s \in S$, then $f \succcurlyeq h$.

To grasp AX, recall that, modulo completeness, $f \succcurlyeq h$ is equivalent to not $h \succ f$. Thus, AX simply says that averaging utilities across states, a.k.a. hedging, cannot make the decision-maker better off, and can make her worse off. A concrete example, pertaining to the particular case where, in AX, f and g are indifferent (i.e., $f \succcurlyeq g$ and $g \succcurlyeq f$), will help seeing how.[18] The example is in terms of the manager betting on the various promotions possible and, for simplicity, with reference to one fixed partition of S. Consider the bets in Example 2, assuming that the payoffs are utility numbers given by the representation of $\widehat{\succcurlyeq}$ (the topic demands cardinal assumptions).

$$\underbrace{\begin{bmatrix} 50 & \text{if } s^* \in E_3 \cap E_1 \\ 10 & \text{if } s^* \in E_3 \cap \overline{E_1} \\ 40 & \text{if } s^* \in \overline{E_3} \cap E_1 \\ 20 & \text{if } s^* \in \overline{E_3} \cap \overline{E_1} \end{bmatrix}}_{f} \sim \underbrace{\begin{bmatrix} 0 & \text{if } s^* \in E_3 \cap E_1 \\ 60 & \text{if } s^* \in E_3 \cap \overline{E_1} \\ 20 & \text{if } s^* \in \overline{E_3} \cap E_1 \\ 30 & \text{if } s^* \in \overline{E_3} \cap \overline{E_1} \end{bmatrix}}_{g} \succ \underbrace{\begin{bmatrix} 25 & \text{if } s^* \in E_3 \cap E_1 \\ 35 & \text{if } s^* \in E_3 \cap \overline{E_1} \\ 30 & \text{if } s^* \in \overline{E_3} \cap E_1 \\ 25 & \text{if } s^* \in \overline{E_3} \cap \overline{E_1} \end{bmatrix}}_{h}.$$

Example 2 – An illustration of AX

By applying first-order stochastic dominance reasoning with reference to (1) and the interpretation of E_1 and E_3 (specified on p. 8), one can check that $f \sim g$. One can also check that in each state, the utility of the payoff of h is half-way between that of f and that of g, i.e., $f(s)h(s)\widehat{\sim}h(s)g(s)$ holds for all $s \in S$. Finally, invoking again first-order stochastic dominance reasoning and (1), one can check that $g \succ h$, thus completing the illustration of AX. In

[17]The techniques applied come from Köbberling and Wakker, 2003. Thus, by building on Alon and Schmeidler's axiomatization, I am invoking techniques which Wakker has championed and he suggests are inapplicable to moral hazard (recall the quotation on p. 4).

[18]In fact, given the other axioms of Alon and Schmeidler's theorem, the indifference restriction of AX (i.e., for any $f, g, h \in F$, if $f \sim g$, and $f(s)h(s)\widehat{\sim}h(s)g(s)$ for all $s \in S$, then $f \succcurlyeq h$) is equivalent to AX itself (see, e.g., Alon, 2015, fn. 6, p. 46, and fn. 8, p. 47).

a nutshell, because she is able to advantageously exert her influence over the events, the decision-maker is unwilling to average utilities across states. With entirely different motivations, Alon and Schmeidler's original "uncertainty aversion" axiom for (2) states exactly the reverse, i.e., the willingness to average utilities across states. To this extent, as previously claimed, AX behaviorally corresponds to a form of "uncertainty seeking".[19] Replacing "uncertainty aversion" with AX in Alon and Schmeidler's theorem yields an axiomatization of (1).[20]

Thus, the most important fact in the present section can be summarized in the following statement. When understood as in (1), *moral hazard is behaviorally indistinguishable from a multi-prior structure coupled with an uncertainty seeking evaluation.*[21] As this well-behaved model is familiar to many, both in economics and in philosophy, this makes transparent that, from an observational and a descriptive point of view, there need to be no incompatibility between moral hazard and the Savage framework.

2 Conceptual discussion of moral hazard

This section turns to a more conceptual discussion of moral hazard. The discussion is motivated by the interaction between moral hazard and the problem of state-dependent utility, which I now introduce.

The problem of state-dependent utility refers to when representations like (1) or (2) must be generalized by replacing the state-*independent* utility function $u : X \to \mathbb{R}$ with a state-*dependent* utility function $u_s : X \times S \to \mathbb{R}$. This is called for in a wide range of circumstances, many of which are of interest regardless of any concern for belief identification. The most intuitive examples are in the economics of health insurance. There, it is essential to allow for the decision-maker's risk attitudes—an aspect of her preferences, reflected in the concavity properties of the representing utility function, and with implications on her demand for insurance—to vary with her state of health and the sub-class of events associated with it (see, e.g., Finkelstein et al., 2013).

[19]This complements a recently proposed interpretation of uncertainty aversion (see Cerreia-Vioglio et al., 2011, Sect. 7). The proposal is that uncertainty aversion can be interpreted as the optimal strategy for a decision-maker playing a game against nature, and nature has the capacity to influence the probability of the events *against the decision-maker's advantage*. Likewise, uncertainty seeking can be interpreted as the optimal strategy for a decision-maker playing a game against nature, and the decision-maker has the capacity to influence the probability of the events *to her own advantage*.

[20]This follows from how Alon and Schmeidler's proof is structured, but it can also be seen directly from the following observation. Take a preference relation $\succcurlyeq$ respecting all the axioms of Alon and Schmeidler, except uncertainty aversion, replaced by AX. Define $\succcurlyeq'$ by $f \succcurlyeq' g$ if and only if $g \succcurlyeq f$. The relation $\succcurlyeq'$ respects all the axioms of Alon and Schmeidler, including uncertainty version. Therefore, $\succcurlyeq'$ is representable as in (2), for some Π and u. By definition, this holds if and only if $\succcurlyeq$ is representable as in (1), for Π and $-u$.

[21]As illustrated in Appendix A.1, this does not hold in general. Specifically, in general, P4 can suffice to distinguish moral hazard from the standard uncertainty seeking model.

Formally, state-dependent utility arises if and only if, for a given decision-maker, a pair of events E, E' can be found, with representing utility functions $u_E, u_{E'}$, such that the utility functions are either not affinely related, or not increasingly related, or not associated with the same range. The first two aspects can be tracked by behavioral regularities such as those described in P4 and P3, respectively (recall the explanations given on p. 9). However, even the conjunction of such regularities does not suffice to behaviorally track the third aspect, while this proves necessary to pin down a unique probability measure. To see this last aspect of the problem, consider, e.g., the traditional subjective expected utility model, i.e., (1) or (2) when Π is a singleton, with $\Pi = \{\pi\}$. For simplicity, assume from now on in this section that the set space is finite, with $S = \{s_1, \ldots, s_n\}$. Observe that, for any full support probability measure $\phi : \Sigma \to \mathbb{R}$, with u_s denoting the collection of u_{s_i} defined by $u_{s_i}(x) = \frac{\pi(s_i)}{\phi(s_i)} u(x)$, for $i = 1, \ldots, n$, we have that, for any $f \in F$:

$$\mathbb{E}^u_\pi(f) = \mathbb{E}^{u_s}_\phi(f). \tag{3}$$

The equality in (3) shows that the decision-maker's betting behavior is represented by u and π if and only if it is represented by u_s and ϕ. Indeed, the traditional uniqueness clause for u in the axiomatizations of the subjective expected utility model must be revised. As (3) shows, even when Savage's state-independent utility axioms hold, u is unique up to a strictly positive affine transformation that can be *state-dependent*. The main implication of this weaker uniqueness class for the utility function is that the probability measure is not, contrary to what is usually taken from the traditional uniqueness clause, absolutely unique. As (3) shows, the probability measure is unique relative to a given transformation of the utility function. The result is that the subjective probability values are, essentially, unidentified.

The above remarks illustrate that when state-dependent utility is not simply excluded ex cathedra (as in the first section of the present paper), the combination of all the possible forms of state-dependent utility creates a conundrum for identifying beliefs based on the observation of betting behavior. This is what I call "the problem of state-dependent utility". The main methodological conclusion from the problem of state-dependent utility is now well established in the literature. It is that even when P3, P4, or the like are respected, and a fortiori when they are not, the betting behavior of a decision-maker does not uniquely identify her underlying beliefs. Therefore, in the terminology of economists, the approach of Savage and followers fails to provide subjective probability measures with "revealed preference" foundations (see especially Drèze, 1987a, Schervish et al., 1990, Karni, 1996, and Baccelli, 2017; see also Dillenberger et al., 2017, Sect. 4.3, for a larger perspective on model misidentification in decision theory). This is a major fact for a philosophical appreciation of the revealed preference methodology.

However, the above conclusion usually comes with the qualification that it admits of one and only one exception: the case of moral hazard.[22] The intuition is that the richer agency of moral hazard situations must leave less degrees of freedom in the representation of betting decisions, thus resolving any remaining undetermination in the identification of the decision-maker's beliefs. Therefore, the complete established view is, in effect, that the betting behavior of a decision-maker uniquely identifies her underlying beliefs *if* (and only if—but I will focus here exclusively on sufficiency) the decision-maker can influence the likelihood of the events upon which she is observed to bet. In the words of Drèze (Drèze, 1987a, p. 29): "'moral hazard' introduces a complication into the Savage model; but that complication is easily handled formally (...); and that complication, in addition to being of independent interest, resolves the identification problem associated with state-dependent preferences." Similarly, Karni systematically opposes two routes which can be taken once the identification issues raised by state-dependent utility are acknowledged (see, e.g., Karni, 2008, p. 228-230). The first is the so-called "hypothetical preference" route of his early work, which achieves identification at the expense of following a choice-based methodology. The second is the moral hazard route of his later work, which would be exceptional in achieving identification following a choice-based methodology.

From now on, within the present paper, call *the sufficiency claim* the claim according to which moral hazard suffices to eschew the identification issues raised by state-dependent utility. A detailed discussion of the results giving credibility to the sufficiency claim in the work of Drèze and Karni is beyond the scope of the present paper. I will focus here, instead, on presenting two self-contained arguments establishing that the sufficiency claim must, in general, be rejected. One argument stems from technical considerations, while the other stems from philosophical considerations.

I discuss the technical considerations first. It is helpful to examine separately each of the two parts of the problem of state-dependent utility, distinguishing them sharply for the sake of the present discussion.

I will be quick on the pure existence side of the problem, a.k.a. the problem of *state-dependent utility with state-dependent preferences*. The problem consists in achieving belief identification when there are underlying pairs $u_E, u_{E'}$ that can be both non-affinely and non-increasingly related. This calls for a joint generalization of both P3 and P4. To say the least, this is challenging. For instance, although P3 and P4 might hold within—but not across—the cells of some partition of the state space, the relevant partitions do not necessarily coincide. This is the case if, e.g., with $S = \{s_1, s_2, s_3, s_4\}$,

[22] Lu's interesting recent work (Lu, 2016) would demand a specific discussion which I cannot provide here. Let me simply remark that Lu's approach crucially relies on uncertainty being *progressively* resolved. This cannot be expressed in the Savage framework. Albeit in different ways and to different degrees, Drèze's and Karni's approaches also rely on uncertainty being progressively resolved (see, e.g., Drèze, 1987a, p. 25, Karni, 2011b, p. 126).

$u_{s_1}(x) = x$, $u_{s_2}(x) = -x$, $u_{s_3}(x) = \sqrt{x}$, and $u_{s_4}(x) = -\sqrt{x}$. None of the rare attempts to generalize both P3 and P4 (see, in particular, Karni, 1992, Karni, 1993, and Hill, 2009) can cope with this simple case. However, given the variety of non-affine transformations possible, it is the simplest of its class. Now, understood as in (1), moral hazard requires P3 and P4 to hold as much as any other decision model (with the qualification given regarding P3 on p. 9). Accordingly, as regards solving the problem of state-dependent utility with state-dependent preferences, moral hazard is, in the Savage framework, in no better position than, say, the traditional subjective expected utility model, i.e., (1) or (2) when Π is a singleton.

What I want to dwell on is, rather, the pure uniqueness side of the problem, a.k.a. the problem of *state-dependent utility without state-dependent preferences*. As (3) illustrates, the problem consists in achieving belief identification when all pairs $u_E, u_{E'}$ are both affinely and increasingly related, but some are not associated with the same range, contrary to what is assumed in the typical state-independent uniqueness clause for u in (1), (2), and the like. Here is the simplest kind of example showing that, understood as in (1), moral hazard does not suffice to solve this problem. Assume that $S = \{s_1, s_2, s_3\}$. Assume further that by applying the variant of Alon and Schmeidler's theorem presented in the previous section, one obtains that the betting behavior of a given decision-maker can be represented as in (1) with the set $\Pi = \{\pi \mid \pi(s_1) = \pi(s_3)\}$ and some state-independent utility function u. Now, reasoning as in (3), observe that the decision-maker's betting behavior is represented by $\Pi = \{\pi \mid \pi(s_1) = \pi(s_3)\}$ and u if and only if it is represented by $\Phi = \{\pi \mid \pi(s_1) = \frac{1}{3}\pi(s_3)\}$ and u_s defined state-wise as follows: $u_{s_1}(x) = 2u(x)$, $u_{s_2}(x) = u(x)$, and $u_{s_3}(x) = \frac{2}{3}u(x)$. This follows from the fact that, for any bet f, and any probability value p,

$$(4) \quad \begin{aligned} & pu\big(f(s_1)\big) + \big(1-2p\big)u\big(f(s_2)\big) + pu\big(f(s_3)\big) \\ &= \frac{1}{2}p \cdot 2u\big(f(s_1)\big) + \big(1-2p\big)u\big(f(s_2)\big) + \frac{3}{2}p \cdot \frac{2}{3}u\big(f(s_3)\big). \end{aligned}$$

The fact that the two representations contrasted above are behaviorally indistinguishable establishes that the problem of state-dependent utility without state-dependent preferences may apply even when the decision-maker can influence the likelihood of the events upon which she is observed to bet.

Drèze anticipated the technical consideration above, but misappreciated it, or so I argue, in one conceptually crucial respect. He (rightfully) remarks that under some assumptions, Π in (1) will prove absolutely unique—i.e., unlike in (4), no alternative state-dependent utility alternative to (1) will be available. This is whenever the stochastic matrix constituted by the elements of Π has full rank, i.e., it has as many linearly independent probability vectors as there are states in the state space (see, e.g., Drèze and Rustichini, 2004, p. 862, Thm. 6.11). Clearly, this condition is not satisfied in the preceding example—witness the linear relationship between $\pi(s_1)$ and $\pi(s_3)$, whether in Π or in Φ. In essence, Drèze interprets this as reinforcing his point that moral hazard is the key to solving the problem of state-dependent utility. He comments on the situation as follows: "the set [Π] is uniquely identified, and so are the units and origins of the utilities, if and only if [Π] is full-dimensional, i.e., if and only if the agent believes that (s)he can influence the probability of every state" (Drèze and Rustichini, 2004, p. 842). But, as the preceding example illustrates, this is partly incorrect. In this example, as no state has a fixed probability value across all the reachable probability measures, the agent *does* believe that she can influence the probability of every state. However, as I have detailed, the set Π is *not* uniquely identified.[23] This means that, in the Savage framework, even making the extremely strong assumption that the decision-maker has full influence over the state space, moral hazard does not, in general, suffice to solve the problem of state-dependent utility.[24]

I now add independent, more philosophical, considerations regarding the sufficiency claim. Whatever the framework in which it is assessed, this claim is, I submit, misleading. This is because even when the betting behavior of the decision-maker can be related to one and only one set Π, this set Π is, by construction, a set of *action-dependent* probabilities.[25] This simple but crucial remark does not apply when Π is arrived at from purely epistemic considerations, as under the usual interpretation of (2). It means that any $\pi \in \Pi$ describes what the decision-maker believes about the state of nature, given one particular way for her to try to exert her influence over it. Whether this dependency can be made explicit, like in Karni's work, or must be left implicit, like in Drèze's work and the present paper, is less important than the fact that it prevails in either case. As a result, the beliefs which the decision-maker holds about the state of nature end up inextricably linked with the beliefs which, rightly or wrongly, she holds about her own agency.

In some cases, namely, when the events make essential reference to the decision-maker's actions (as with E_1 introduced in the previous section, on p. 8), the link between beliefs and agency is literally inextricable. But even

[23]If a row-stochastic matrix has full rank, then it has no constant column, but the converse does not hold. This invalidates the "if" direction of the last part of Drèze's comment.

[24]Notice that, under the assumptions of the present paper, even observing the actions might not lead to any progress in identification. Enriching the observable choice set from F to $F \times A$ could give principled grounds to couple each $\pi \in \Pi$ with some $a \in A$, and lead to the construction of a set of $\pi_a \in \Pi \times A$. But the identification issues illustrated by (4) apply with respect to a set of $\pi_a \in \Pi \times A$, no less than a set of $\pi \in \Pi$.

[25]Action-dependent probabilities are not probabilities about one's actions. Thus neither Drèze or Karni, nor myself need to disagree with the philosophical claim according to which "[a]ny adequate (...) decision model must not explicitly or implicitly contain any subjective probabilities for acts" (Spohn, 1977, p. 114; see also Levi, 1993 and Gaifman, 1999). The distinction between control and influence helps in further demystifying action-dependent probabilities, because skeptics tend to exclusively focus on cases of full control.

when this link is in principle extricable (as with E_2 in the previous section), there is no reason to assume that the two components can actually be disentangled. Such would be the case if, from the set Π, one had a principled way of picking or, alternatively, constructing one and only one distinguished probability measure π^* that would capture the decision-maker's *action-independent* beliefs about the state of nature.[26] At least when the influencing actions are not observable, as in the present paper and the standard understanding of the problem of moral hazard, there is no reason to hope that this new identification problem can be solved. Therefore, at the end of the day, one has essentially traded one identification problem for another.

However, it cannot be emphasized enough that the problem of state-dependent utility would be genuinely solved only if this new identification problem could be solved. This is because there is no doubt that the initial identification program pertains to action-independent beliefs. Otherwise, some of the motivations for identifying beliefs rather than merely predicting behavior (as advocated in, e.g., Nau, 2001) would be undermined. Such is the case, for instance, of the simple but key idea according to which the decision-maker's beliefs can be used in someone else's decision-making, e.g., the observer's (see, e.g., Drèze and Rustichini, 2004, p. 848-849). I therefore conclude that, in the present state of decision theory, where no moral hazard identification strategy leads to a unique action-independent probability measure, it is misleading to present moral hazard as offering a solution to the identification issues raised by state-dependent utility.

Thus, Drèze may not be pointing to the most significant limitation of the approach which he pioneered when he (rightfully) highlights (Drèze and Rustichini, 2004, p. 848): "[i]ts realm of application remains limited, of course, by the extent to which states are subject to moral hazard. For states lying entirely outside the [influence] of the agent, like the weather or macroeconomic realisations, this approach is of no use."[27] Even when all the states are under the influence of the decision-maker, so that perhaps *some identification* can be achieved, there remains the equally important question of understanding *what, really, is identified*. Because of the very nature of the proposed identification strategy, this has become obscure along the way.

[26] The representation in Karni, 2011b features an action-independent probability measure (see Eq. 2, p. 129). But it is over some signals which the decision-maker receives in a first step, not the events which she can influence in a second step. The representation in Karni, 2011a also features an action-independent probability measure, with deeper foundations. This important paper would deserve a specific discussion, which I cannot provide here. I can only highlight that it crucially relies on a profound transformation of the very notion of a state space and, thereby, of what "action-independence" can mean.

[27] For terminological consistency, I have substituted "influence" for "control" in original.

Conclusion

In this paper, I have investigated the betting behavior of a decision-maker who has the capacity—or so she thinks—to affect the likelihood of the events upon which she is offered to bet. In decision theory, this is best known as a case of moral hazard. Such cases have been the object of some interest in themselves, as well as in connection with the identification issues raised by state-dependent utility.

First, I have sketched the first systematic axiomatic analysis of moral hazard in Savage's purely subjective framework. I have shown that a particularly simple form of moral hazard amounts to a multi-prior structure coupled with an uncertainty seeking evaluation. This makes transparent that, from an observational and a descriptive point of view, there need to be no incompatibility between moral hazard and the Savage framework. This importantly qualifies the incompatibility view that is ubiquitous in decision theory, especially in the causal decision theory literature, where it seems to be held across a wide spectrum of otherwise conflicting philosophical positions. By and large, the pre-existing results of Drèze and Karni could have sufficed to suggest that the incompatibility view calls for qualifications. But this is now established in all necessary details, thanks to the dedicated analysis which I have provided in the Savage framework.

Second, I have offered grounds to reject another influential view, endorsed by Drèze, Karni, and fellow economists, to the effect that moral hazard suffices to eschew the identifications issues raised by state-dependent utility. Still based on the simple form of moral hazard above, I have shown that these issues can arise in the Savage framework even when the decision-maker has the capacity to affect the likelihood of the events of interest—indeed, even when she has such capacity over all the events of interest. I have also argued that, because moral hazard situations lead to the elicitation of beliefs that are about both the state of nature and the decision-maker's assessment of her potential influence over it, it is philosophically misguided to hope that they can genuinely solve the identification issues raised by state-dependent utility. In lieu of a solution, this amounts to trading an identification problem for another, with no compelling progress along the way.

Appendix A.1

Take E_1 and E_3, as introduced on p. 8. E_1 is under the control of the decision maker. Departing from the assumption that actions do not have variable intrinsic utilities, assume now that this control is exerted thanks to some action a, the disutility of which increases with the probability of E_1, $\pi(E_1)$. For simplicity, assume that this disutility is separable from the utility of the payoffs, with, say, $u(a) = -20\pi(E_1)$. By contrast with E_1, E_3 is not under the influence of the decision-maker, so that no utility cost is associated with it. Finally, assume that the payoffs of the bets introduced below are utility numbers, in the sense that the decision-maker has a linear utility function for money. (The topic demands cardinal assumptions.) Then, one can check that the decision-maker will make the following betting decisions, which taken together constitute a violation of P4: $100 \succ 0$, $50 \succ 40$, $100E_1 0 \succ 100E_3 0$, and $50E_1 40 \prec 50E_3 40$. This proves that not all forms of moral hazard respect P4.

Appendix A.2

This appendix lists the axioms which Alon and Schmeidler used to provide an axiomatization for (2) (see Alon and Schmeidler, 2014).[28]

A0 S is non-empty, X is a connected topological space, and F is endowed with the product topology.

A1 $\succcurlyeq$ is transitive and complete.

A2 $\succcurlyeq$ is continuous, i.e., for all $f \in F$, the sets $\{g \in F \mid g \succ f\}$ and $\{g \in F \mid g \prec f\}$ are open in the product topology.

A3 There is an essential event, i.e., there exists E $\in \Sigma$, and $x, y \in X$, such that $x \succ xEy \succ y$.

A4 $\succcurlyeq$ respects state-wise dominance, i.e., for all $f, g \in F$, if $f(s) \succcurlyeq g(s)$ for all $s \in S$, then $f \succcurlyeq g$.

A5 $\succcurlyeq$ respects binary co-monotonic trade-off consistency, i.e., for all $a, b, c, d, w, x, y, z \in X$, and $E, E' \in \Sigma$, if $aEw \sim bEx$ and $cEw \sim dEx$, and $aE'y \sim bE'z$, then $cE'y \sim dE'z$, $\{aEw, bEx, cEw, dEx\}$ and $\{aE'y, bE'z, cE'y, dE'z\}$ being sets of co-monotonic bets, E being non-null on the maximal co-monotononic extension of the first set, E', on that of the second.

A6 $\succcurlyeq$ displays uncertainty aversion, i.e., for all $f, g, h \in F$, if $f \succcurlyeq g$, and $f(s)h(s) \widehat{\sim} h(s)g(s)$ for all $s \in S$, then $h \succcurlyeq g$.

A7 $\succcurlyeq$ respects certainty independence, i.e., for all $f, g \in F$, and $w, x, y \in X$, if $wx \widehat{\sim} xy$, and $f(s)g(s) \widehat{\sim} g(s)y$ for all $s \in S$, then $g \sim x$ if and only if $f \sim w$.

A8 $\succcurlyeq$ respects certainty co-variance, i.e., for all $f, g \in F$, and $x, y \in X$, if $f(s)g(s) \widehat{\sim} xy$ for all $s \in S$, then $f \sim x$ if and only if $g \sim y$.

[28] Recall that the abbreviation $\widehat{\sim}$ has been introduced on p. 11.

References

Alon, S. (2015): "Worst-Case Expected Utility", *Journal of Mathematical Economics*, 60, 43–48.

Alon, S., and D. Schmeidler (2014): "Purely Subjective Maxmin Expected Utility", *Journal of Economic Theory*, 152, 382–412.

Anscombe, F., and R. Aumann (1963): "A Definition of Subjective Probability", *The Annals of Mathematical Statistics*, 34(1), 199–205.

Baccelli, J. (2017): "Do Bets Reveal Beliefs?", *Synthese*, 194(9), 3393–3419.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011): "Uncertainty Averse Preferences", *Journal of Economic Theory*, 146(4), 1275–1330.

Dillenberger, D., A. Postlewaite, and K. Rozen (2017): "Optimism and Pessimism with Expected Utility", *Journal of the European Economic Association*, 15(5), 1158–1175.

Drèze, J. (1961): "Les fondements logiques de la probabilité subjective et de l'utilité", *La décision. Colloques Internationaux du Centre National de la Recherche Scientifique*, pp. 73–87.

——— (1987a): "Decision Theory with Moral Hazard and State-Dependent Preferences", in *Essays on Economic Decisions under Uncertainty*, ed. by J. Drèze, pp. 23–89. Cambridge: Cambridge University Press.

——— (1987b): "Logical Foundations of Cardinal Utility and Subjective Probability", in *Essays on Economic Decisions under Uncertainty*, ed. by J. Drèze, pp. 90–104. Cambridge: Cambridge University Press.

Drèze, J., and A. Rustichini (1999): "Moral Hazard and Conditional Preferences", *Journal of Mathematical Economics*, 31(2), 159–181.

——— (2004): "State-Dependent Utility Theory", in *Handbook of Utility Theory, Volume II: Extensions*, ed. by S. Barbera, P. Hammond, and C. Seidl, pp. 839–892. Boston: Kluwer Academic Press.

Ellsberg, D. (1961): "Risk, Ambiguity, and the Savage Axioms", *The Quarterly Journal of Economics*, 75(4), 643–669.

Epstein, L., and M. Le Breton (1993): "Dynamically Consistent Beliefs Must Be Bayesian", *Journal of Economic Theory*, 61(1), 1–22.

Finkelstein, A., E. Luttmer, and M. Notowidigdo (2013): "What Good is Wealth without Health? The Effect of Health on the Marginal Utility of Consumption", *Journal of the European Economic Association*, 11(1), 221–258.

Fishburn, P. (1970): *Utility Theory for Decision Making*. New York: Wiley.

Gaifman, H. (1999): "Self-Reference and the Acyclicity of Rational Choice", *Annals of Pure and Applied Logic*, 96(1-3), 117–140.

Ghirardato, P., and M. Marinacci (2001): "Risk, Ambiguity, and the Separation of Utility and Beliefs", *Mathematics of Operations Research*, 26(4), 864–890.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non-Unique Prior", *Journal of Mathematical Economics*, 18(2), 141–153.

Hart, O., and B. Hölmstrom (1987): "The Theory of Contracts", in *Advances in Economic Theory*, ed. by T. Bewley, pp. 71–155. Cambridge: Cambridge University Press.

Hart, S., S. Modica, and D. Schmeidler (1994): "A Neo^2 Bayesian Foundation of the Maxmin Value for Two-Person Zero-Sum Games", *International Journal of Game Theory*, 23(4), 347–358.

Hill, B. (2009): "When Is There State Independence?", *Journal of Economic Theory*, 144(3), 1119–1134.

Jeffrey, R. (1965): *The Logic of Decision*. New York: McGraw-Hill Book Company (First Edition).

——— (1983): *The Logic of Decision*. Chicago: Chicago University Press (Second Edition).

Joyce, J. (1999): *The Foundations of Causal Decision Theory*. New York: Cambridge University Press.

Karni, E. (1992): "Subjective Probabilities and Utility with Event-Dependent Preferences", *Journal of Risk and Uncertainty*, 5(2), 107–125.

——— (1993): "Subjective Expected Utility Theory with State-Dependent Preferences", *Journal of Economic Theory*, 60(2), 428–438.

——— (1996): "Probabilities and Beliefs", *Journal of Risk and Uncertainty*, 13(3), 249–262.

——— (2006): "Subjective Expected Utility Theory without States of the World", *Journal of Mathematical Economics*, 42(3), 325–342.

——— (2008): "State-Dependent Utility", in *The Handbook of Rational and Social Choice*, ed. by P. Anand, P. Pattanaik, and C. Puppe, pp. 223–238. Oxford: Oxford University Press.

——— (2011a): "Subjective Probabilities on a State Space", *American Economic Journal: Microeconomics*, 3(4), 172–185.

——— (2011b): "A Theory of Bayesian Decision Making with Action-Dependent Subjective Probabilities", *Economic Theory*, 48(1), 125–146.

——— (2013): "Bayesian Decision Theory with Action-Dependent Probabilities and Risk Attitudes", *Economic Theory*, 53(2), 335–356.

Köbberling, V., and P. Wakker (2003): "Preference Foundations for Nonexpected Utility: A Generalized and Simplified Technique", *Mathematics of Operations Research*, 28(3), 395–423.

Levi, I. (1993): "Rationality, Prediction, and Autonomous Choice", *Canadian Journal of Philosophy*, 23(suppl.), 339–363.

Lu, J. (2016): "A Bayesian Theory of State-Dependent Utilities", working paper, accessible online at the following address: http://www.econ.ucla.edu/jay/bayes.pdf.

Nau, R. (2001): "de Finetti Was Right: Probability Does Not Exist", *Theory and Decision*, 51(2-4), 89–124.

Savage, L. (1954): *The Foundations of Statistics*. New York: Wiley (First Edition).

——— (1972): *The Foundations of Statistics*. New York: Dover (Second Edition).

Schervish, M., T. Seidenfeld, and J. Kadane (1990): "State-Dependent Utilities", *Journal of the American Statistical Association*, 85(411), 840–847.

Spohn, W. (1977): "Where Luce and Krantz Do Really Generalize Savage's Decision Model", *Erkenntnis*, 11(1), 113–134.

Wakker, P. (2010): *Prospect Theory–For Risk and Ambiguity*. Cambridge: Cambridge University Press.