CAN REDESCRIPTIONS OF OUTCOMES SALVAGE THE AXIOMS OF DECISION THEORY?[a]

Jean Baccelli[*] and Philippe Mongin[**]

16 July 2021

Abstract. The basic axioms or formal conditions of decision theory, especially the ordering condition put on preferences and the axioms underlying the expected utility (EU) formula, are subject to a number of counter-examples, some of which can be endowed with normative value and thus fall within the ambit of a philosophical reflection on practical rationality. Against such counter-examples, a defensive strategy has been developed which consists in redescribing the outcomes of the available options in such a way that the threatened axioms or conditions continue to hold. We examine how this strategy performs in three major cases: Sen's counterexamples to the binariness property of preferences, the Allais paradox of EU theory under risk, and the Ellsberg paradox of EU theory under uncertainty. We find that the strategy typically proves to be lacking in several major respects, suffering from logical triviality, incompleteness, and theoretical insularity (i.e., being cut off from the methods and results of decision theory). To give the strategy more structure, philosophers have developed "principles of individuation"; but we observe that these do not address the aforementioned defects. Instead, we propose the method of checking whether the strategy can overcome its typical defects once it is given a proper *theoretical expansion* (i.e., it is duly developed using the available tools of decision theory). We find that the strategy passes the test imperfectly in Sen's case and not at all in Allais's. In Ellsberg's case, however, it comes close to meeting our requirement. But even the analysis of this more promising application suggests that the strategy ought to address the decision problem as a whole, rather than just the outcomes, and that it should extend its revision process to the very statements it is meant to protect. Thus, by and large, the same cautionary tale against redescription practices runs through the analysis of all three cases. A more general lesson, simply put, is that there is no easy way out from the paradoxes of decision theory.

Keywords: decision theory, binariness property, expected utility, Sen, von Neumann and Morgenstern, Savage, Allais paradox, Ellsberg paradox, redescription of outcomes, principles of individuation.

---

[a] For helpful comments and suggestions, the authors are grateful to two reviewers, Associate Editor Blake Roeber, and the 2020 Spring Fellows at the Center for Philosophy of Science, University of Pittsburgh.

[*] Munich Center for Mathematical Philosophy. Email: jean.baccelli@gmail.com (corresponding author)

[**] CNRS & HEC Paris; LEMMA, Université de Paris 2 Panthéon Assas.
Email: mongin@greg-hec.com

1. Introduction

The axiomatic conditions of decision theory face counter-examples in the form of idealized choice problems in which they appear to be defeated. Famously, the expected utility (EU) formula and its underlying axioms, whether in von Neumann-Morgenstern's (1944-1947) probabilistic risk version or in Savage's (1954-1972) pure uncertainty variant, have attracted such counter-examples – most notably the Allais and Ellsberg paradoxes. Even such a basic condition as the transitivity of preferences has fostered a plethora of counter-examples, among which Sen's (1993) are perhaps the best known. Among decision theorists, a widespread reaction has been to deny these counter-examples normative relevance, while possibly granting them empirical relevance. The claim is then simply that the problematic choices are not rationally defensible, despite being choices which real agents possibly could make. However, this dichotomous strategy is not easily applicable, because the best-known counter-examples are ambiguous between the normative and the positive, and provide fuel for our reflection on the norms of practical rationality. This, indeed, is why they can be properly considered *paradoxes*, displaying a direct conflict between the intuition that the idealized choices presented are rational, and the intuition that the axiomatic conditions these choices contravene are principles of rationality. As a result, an alternative strategy has developed, primarily among philosophers of decision theory, which consists in *redescribing* the choice problems of the counter-examples so that the choices made no longer violate the axioms of decision theory. In other words, the strategy is to *redescribe paradoxes away*, i.e., to dissipate through redescription any conflict between our normative intuitions. Proponents of this strategy usually target their redescriptions at the *outcomes* of the available options, rather than at other components of the choice problems.[1] Of course, any such changes would need to be justified, but the authors in question have shown some ingenuity in making them plausible.

The present paper aims at determining whether the redescription strategy, in its common redescription-of-outcomes form, provides the axiomatic conditions with the degree of protection that its proponents claim. As the strategy has emerged in a normative context of evaluation, we will interpret it accordingly, seeking thus to determine whether it actually

[1] Savage (1954-1972) uses the word "consequence" to refer to what results from an uncertain prospect in a given state; Luce and Raiffa (1957) use "outcome" to refer to what results from a lottery when this lottery is drawn. This is a reasonably well-established distinction; but we will use the word "outcome" throughout, both for simplicity and to follow the antecedent literature on redescription.

supports the claim that the counter-examples do not threaten the rationality status of the axiomatic conditions, and, more generally, to assess whether it should have the place it has in the theoretical work on practical rationality. Although we grant that the strategy might also offer resources to those who deny that the counter-examples are proper empirical refutations, we will not explore that part of the picture here. Even restricted to the normative interpretations, the strategy has a sufficiently large range of occurrences to keep the philosopher of decision theory busy. We illustrate this range by selecting for examination three theories of graded complexity with accompanying counter-examples. In the context of certainty, we look at the theory of choice functions and the binariness property of preference, which is a more primitive condition than transitivity; the counter-examples are Sen's. In the context of probabilistic risk, we look at von Neumann and Morgenstern's brand of EU theory, with its standard counter-example, the Allais paradox. In the context of pure uncertainty, we look at Savage's brand of EU theory, again with its standard counter-example, the Ellsberg paradox. These three counter-examples and the accompanying redescriptions comprise a limited sample; yet this sample is representative of the most fundamental discussions currently ongoing in normative decision theory.

The redescription-of-outcomes strategy is sometimes employed in decision theory itself, yet it remains elusive there: rarely made fully explicit, let alone theorized. There is a contrast here with the more reflective work that has been pursued by philosophers: Broome (1991), Schick (1991), Dreier (1996), Bermudez (2007), Buchak (2013), and Bradley (2016, 2017) have thoroughly discussed it.[2] These authors usually approach the strategy from the angle of providing "principles of individuation" that are meant to assign redescription its proper domain of application. Our work is in part a reaction to theirs, and we will accordingly also discuss these principles briefly. But we respond here mainly as decision theorists, concerned as we are that redescription practices block decision theory from playing its proper role in the work on practical rationality. In fact, our main argument follows a new direction, orthogonal to the one followed by the principles of individuation, and which we now explain.

Existing redescriptions have striking defects that our sample will help make precise: they only deliver logically weak conclusions; they are often carried out incompletely; and

[2] The redescription-of-outcomes strategy has sometimes attracted philosophers' attention for reasons other than the defence of axiomatic conditions against counter-examples. Thus, Weirich (1986 and forthcoming) argues that a proper theory of risk and risk-attitude should include the risk of a lottery in the description of its outcomes. Following a different line, Verbeek (2001) discusses the proper definition of outcomes in connection with the decision-theoretic principle of consequentialism, a principle defended by Hammond (1988) and rejected by Machina (1990). Consequentialism is an important topic both for decision theory and its philosophical appraisal; but we will not discuss it here.

they are theoretically insular, in the sense that they are cut off from the methods and results of decision theory. To remedy the last defect may also be a way of remedying the first two, so we lay emphasis on the use of what we call *theoretical expansion* – i.e., using the tools of decision theory to develop the proposed redescription – as a touchstone to evaluate each application of the strategy. At this stage of our argument, then, we part company with the earlier literature, whose "principles of individuation" appear to be of limited use since they apply to redescriptions simply as they are, i.e., with the defects listed at the beginning of this paragraph. As we will argue, our criterion of demanding a compelling theoretical expansion does a better job than the principles of individuation at identifying which redescriptions may be well-founded and which are not. In accordance with our touchstone, we expand theoretically on the three cases of our sample, a task that unsurprisingly grows in complexity when one goes from the context of certainty to those of probabilistic risk and then of pure uncertainty. This process will be shown to lead to variable results. Using some sophisticated work on choice functions, we will be able to conclude that binariness and related consistency conditions can be salvaged from Sen's counter-examples, though still only trivially. Concerning the Allais paradox, from the toolbox of available decision theories we use those of regret and disappointment avoidance. This leads to a dilemma that some may find surprising: either the theoretical expansion preserves the endangered EU axiom, or it reproduces the Allais paradox, but it cannot achieve both goals at once. As to the Ellsberg paradox, we relate the theoretical expansion to second-order expected utility theory, a generalization of EU theory in Savage's version that connects well with the chosen redescription strategy. All in all, this last example will emerge as the most satisfactory of the lot, which may be explained by the fact that the strategy here is not narrowly focused on outcomes, but also involves a careful restatement of the states of the world. That our conclusions are not uniform is in line with the nature of our main recommendation, which entails assessing each case on the merits of its particular theoretical development. Just as we do not want to presume of any proposed redescription that it is justified or successful, we do not want to presume that none could ever have these properties. Emphatically, this is a case-by-case inquiry. That being said, a general lesson emerges from the various cases we examine, namely that there is no point trying to avoid the paradoxes of decision theory by redescribing them away; these paradoxes must be faced squarely. Just as Sen, Allais, or Ellsberg would readily accept, sooner or later one must pick one's camp between defending the rationality of the principles under attack, or defending the rationality of the contravening preferences.

The paper is organized as follows. Sections 2 and 3 provide a succinct statement of Sen's and Allais's counterexamples respectively, along with the redescription strategies that have been proposed as salvaging devices. These two cases provide a sufficient starting point, and the study of Ellsberg's counterexample is deferred to a later part of the paper. Section 4 offers a brief critical review of the "principles of individuation" currently under debate in the philosophical literature. Section 5 points out that they do not respond to the main objections, which are listed there in detail, and it introduces our solution in terms of theoretical expansion. Sections 6 and 7 evaluate the redescriptions in Sen's and Allais's cases, based on their respective theoretical expansions. Section 8 deals with the Ellsberg paradox and the strategies employed to overcome it. Section 9 concludes.

2. Binariness and its failures

Two very simple and well-known examples provided by Sen (1993, p. 501-502) comprise a touchstone for the redescription-of-outcomes strategy in the context of decision under certainty. Here the reformulation bears on the *options*, but it can equally be said to bear on the *outcomes*, since this context of decision precludes any distinction between these two concepts. In the first example, two agents are presented with two different sets of pieces of cakes. In one set there are three pieces x, y, and z, the sizes of which increase in that order, and in the other set there are only the two smaller pieces x and y. The agent who chooses first takes y from the former set and x from the latter. This violates the contraction property, also known as the Chernoff condition or Sen's "property α".[3] This property requires in particular that if y is chosen from the option set $\{x, y, z\}$, then y should still be chosen when the option set is contracted to $\{x, y\}$. This is necessary for a choice function to be *binary*, i.e., to originate in the maximization of a binary relation. This last requirement is even more basic than transitivity, since transitivity can only hold of a well-defined binary relation.[4]

[3] Formally, let us denote the choice function by C(.) and two subsets of options by A, B. Property α then says that if $B \subset A$ and $x \in C(A)$, then $x \in C(B)$.

[4] As Sen (1982, ch. 1) explains, when the choice function is singleton-valued, property α is necessary and sufficient for the choice function to originate in an ordering; generally, though, it is only necessary. The definition of binariness adopted in this work by Sen, and taken up here, puts no restriction on the binary relation in the maximization of which the choice function originates. Elsewhere, Sen (1993, p. 500) considers a definition that requires the choice function to originate in the maximization of a particular relation (the "revealed preference" relation). *A fortiori*, violations of property α contradict this stronger sense of binariness.

In Sen's second example, the agent is first offered the choice between going to an acquaintance's home and having tea (x), and not going there (y), and she chooses x over y. Then she is given the choice between the same x and y, and going to the acquaintance's home and having cocaine instead of tea (z), and she chooses y. This again violates the contraction property, so that the agent cannot be endowed with a binary relation she would maximize, let alone satisfy transitivity.[5]

Sen offers psychological explanations for these examples: in the first, the agent appears to want to be polite by always taking the second-largest piece of cake, and in the second, she appears to be frightened by what she has discovered about her acquaintance's lifestyle. Heuristically speaking, in both examples the opportunity set determines not only the physical availability of choices, but also the meanings attached to these choices, a phenomenon technically referred to in the literature as *menu-dependence*. Each example illustrates this phenomenon in a different way. In the first, the agent intends to follow a rule of conduct and the opportunity set indicates what obeying this norm means given the physical possibilities. In the second, the agent lacks a piece of information that would change the meaning she gives to her choices, and the second opportunity set brings her this information, thus resolving the uncertainty attached to the first opportunity set.[6]

Although this is not Sen's point (what he wants to demonstrate is that "internal", i.e., purely formal, choice conditions do not exhaust the rationality concept), it is easy to connect his examples with the redescription-of-outcomes strategy. Arguably, the first agent is not choosing from $\{x, y\}$ and $\{x, y, z\}$, but rather from

> $\{x$ for her and y for the other, y for her and x for the other$\}$ and $\{x$ for her and y or z for the other, y for her and x or z for the other, z for her and x or y for the other$\}$.

Another, more drastic redescription would be:

> {next-to-largest piece, largest piece} and {smallest piece, next-to-largest piece, largest}.[7]

Similarly, the frightened agent would be choosing from

> {not going, going and having tea when tea is the only offer} and {not going, going and having tea when tea and cocaine are offered, going and having cocaine when tea and cocaine are offered}.

---

[5] A famous early variant of Sen's second example appears in Luce and Raiffa (1957, p. 288).
[6] For more comments, see Dhongde and Pattanaik (2010, p. 22) and Battacharyya, Pattanaik and Xu (2011, p. 127-128). The latter writers neatly separate "menu-dependent criteria", as in the first Sen example, and "menu-dependent information", as in the second.
[7] Suggested by Neumann (2007).

With these reformulations, no set-theoretic comparison between the choice sets is possible anymore, and the contraction property vacuously holds. This is meant to salvage the property in question.

3. Expected utility and the Allais paradox

Allais's (1953 and 1979) paradox is a thought experiment that challenges the brand of decision theory developed by von Neumann and Morgenstern (1944-1947) and their followers – VNM theory for short. In this particular form of EU theory, the objects of preference are *lotteries*, which are defined as probability measures on some set representing the final outcomes. In the axiom system, the prominent condition is *VNM independence*, which says in effect that if a lottery p is preferred to a lottery q, this preference is unchanged when they are mixed with a third lottery r in the same proportion. The VNM representation theorem establishes that a preference relation satisfying the axioms can be represented by an EU formula adapted to the lottery setting. Allais directed his paradox against both this representation in general and VNM independence specifically. While later decision theorists have turned the paradox into an empirical refutation, Mongin (2019) emphasizes that Allais meant it to be a normative argument, and proposes to reinterpret it accordingly. We follow the same line here.[8]

Allais posits an ideal agent who is asked to make two choices in succession, first between lotteries p_1 and q_1, and then between lotteries p_2 and q_2 (the numbers are million French francs):

p_1 : 100 with prob 1	p_2 : 100 with prob 0.11; 0 with prob 0.89
q_1 : 500 with prob 0.10; 100 with prob 0.89; 0 with prob 0.01	q_2 : 500 with prob 0.10; 0 with prob 0.90

Table 1

Two pairs of choices, (p_1, p_2) and (q_1, q_2), are compatible with an EU representation, while the other two pairs, (p_1, q_2) and (q_1, p_2), are not. This is readily checked by noting that, for any

[8] Mongin (2019) also covers some of the vast secondary literature.

utility function u defined on the monetary outcomes, the following two inequalities are equivalent:

$$u(100) > 0.10\, u(500) + 0.01\, u(0) + 0.89\, u(100)$$

and

$$0.11\, u(100) + 0.89\, u(0) > 0.10\, u(500) + 0.90\, u(0).$$

Allais's proposed resolution of the two choice problems is (p_1, q_2), which thus violates the conclusion of the VNM representation theorem.[9]

Allais gives two explanations for why this pair might be reasonable. On the one hand, he stressed that, in the first choice situation, given the relatively high gain it secures, the absolute certainty of p_1 suggests that one should choose it, however small the chance of getting nothing in the alternative q_1. Arguably, such a prudent attitude would be compatible, in the second choice situation, with p_2 being chosen over q_2, i.e., with the riskier of the two lotteries being chosen, because in this case the chances of getting nothing are nearly equal whereas the possible gains are substantially different. On the other hand, more abstractly (and perhaps more obscurely), Allais argued that the 1% chance of getting 0 in lottery q_1 in the first situation reappeared, in the second, as a difference between the 90% chance of getting nothing in q_2 and the 89% chance of getting 0 in p_2. As he went on to argue, this same 1% chance had a very different psychological effect in the two choice situations. Being isolated in the first, it had a weight it could not have in the second, where it is surrounded by other chances of getting 0. We will return to this *complementarity* argument, which goes more deeply into the sense of the paradox than the *certainty* argument.[10]

A common way of discussing the Allais paradox is to suppose that the two choice situations are implemented by a concrete randomization device, for instance an urn with 100 numbered balls that are mapped onto monetary outcomes as follows:

[9] To see that this resolution violates VNM independence more specifically, define the auxiliary lottery l giving outcome 500 with probability 10/11 and outcome 0 with probability 1/11. Under the ordering condition, by VNM independence, choosing p_1 over q_1 implies choosing 100 over l, while choosing q_2 over p_2 implies choosing l over 100 – a contradiction.

[10] This may not be Allais's own evaluation of the two arguments; but it certainly is that of Machina (1987) and other technical followers of Allais.

	Balls 1-10	Ball 11	Balls 12-100
p^*_1	100	100	100
q^*_1	500	0	100
p^*_2	100	100	0
q^*_2	500	0	0

Table 2

Although this restatement seems innocuous, it is technically appropriate neither for introducing the Allais paradox, nor for deciding whether it achieves its theoretical aim. By definition, lotteries attribute probabilities directly to final outcomes without referring to states of the world. But Table 2 refers to states of the world, represented therein by ball numbers. This amounts to replacing the VNM objects of preference by Savage's (1954-1972) different objects, which are *uncertain prospects*, as defined by mappings from states to outcomes. With this surreptitious change, the focus of attention shifts from VNM independence to Savage's *sure-thing principle* (STP). Although the two conditions superficially resemble each other, they are mathematically distinct. Roughly speaking, there is more to VNM independence than the STP. In any case, salvaging the latter would not yet leave the former on safe ground. However, the amalgamation of the two conditions is so entrenched in the philosophical literature that we let it pass, and will in our turn discuss the Allais paradox in the reformulation of Table 2 as if it were still concerned with the STP.[11]

This principle, which appears as postulate P2 in Savage's system, says in essence that if a preference comparison holds between two uncertain prospects f and g, which have identical consequences over some part P of the state space, this preference comparison remains unchanged when the identical consequences of f and g on P receive a new set of values, and these two prospects remain otherwise the same. In the particular instance, f and g are p^*_1 and q^*_1, P is the event "Balls 12-100", and 100 is the initial common value on P. The changes of p^*_1 to p^*_2, and of q^*_1 to q^*_2, amount to bringing this common value to 0, while keeping everything else the same. By the sure-thing principle, this change should be immaterial, i.e., one should have that p^*_1 is preferred to q^*_1 if and only if p^*_2 is preferred to q^*_2; however, the choice (p^*_2, q^*_2) does violate this equivalence.

[11] After Broome (1991) and Schick (1991), Bermudez (2009), Buchak (2013), Bradley (2017) and Bradley and Stefansson (2017) have discussed the Allais paradox in terms of the STP.

We now report a redescription of outcomes that has been proposed to deflect this conclusion. Allegedly, q^*_1 or q^*_2 do not have the same outcome on Ball 11. If the agent has chosen q^*_1 and Ball 11 is drawn, she is most likely to regret not having chosen p^*_1. But if she has chosen q^*_1 and Ball 11 is drawn, the comparison she will make with the unchosen alternative, here p^*_2, is counterbalanced by the thought that the unrealized states Balls 12-100 would have brought about the same result, so she is not likely to feel regret to the same extent as with q^*_1. In other words, the fact that the 0 monetary amount is isolated in one case and spread out in the other makes a difference to the agent's negative feelings. The redescription strategy jumps from this psychological argument (which is not unlike Allais's) to the conclusion (which would not be Allais's) that outcomes should be relabelled as in the next table.

	Balls 1-10	Ball 11	Balls 12-100
p^*_1	100	100	100
q^*_1	500	0 with regret	100
p^*_2	100	100	0
q^*_2	500	0 without regret	0

Table 3

A variant exists that mentions "disappointment" instead of "regret". Either way, if the STP is to apply, it must be the case that p^*_1 be equal to p^*_2, and q^*_1 equal to q^*_2, outside the P subset of states of the world where p^*_1 and q^*_1, and p^*_2 and q^*_2, respectively coincide. However, with the proposed redescription, this condition fails regarding q^*_1 and q^*_2. Thus, in Table 3, the STP holds vacuously, instead of being violated as it was in Table 2.[12]

4. A common objection and a response in terms of "principles of individuation"

An obvious worry, which has surfaced even among proponents of the redescription-of-outcomes strategy, is that this threatens to empty decision theory of its content. The worry

[12] This pattern of response to the Allais and related paradoxes goes back as far as Raiffa (1968, p. 85-86), who, however, did not mean to endorse it. It has occasionally reappeared in later decision theory, albeit allusively. It has been actively discussed by philosophers; see the references in fn 11.

was expressed long ago among decision theorists who witnessed the first counter-examples to EU theory. Thus, in connection with the Allais paradox, Raiffa (1968, p. 86) complained that the "flexibility" of the strategy "has been bought at a price: no longer is zero the same as zero…". In connection with one of Savage's postulates, P4, MacCrimmon and Larsson wrote: "If we were to always redefine the consequences whenever [P4] was apparently violated, then [P4] would be tautological. Thus there must be some balance struck between redefining consequences to avoid a violation and letting a violation stand" (1979, p. 397-398). Similarly, after pondering over the Allais paradox and a response to it in terms of regret, Tversky wrote: "In the absence of any constraints, the consequences can always be interpreted so as to satisfy the axioms. In this case, however, the theory becomes empty *from both descriptive and normative standpoints*" (1975, p. 171, our emphasis). As the end of the last quotation makes clear, the redescription-of-outcomes strategy threatens to undermine not only the empirical testability of decision theory, but also its normative function as a provider of principles of rational action.

The philosophy of decision theory has attempted to address this nagging problem by placing restrictions on the scope of the strategy. All of these proposals take the form of "principles of individuation", the practical role of which is to tell the decision theorist when he is entitled to depart from the *prima facie* descriptions of the outcomes, and when he is not. As the strategy supposedly brings to light different descriptions that were initially conflated, these principles can be viewed more specifically as putting checks on possible refinements.[13] In the present section we briefly discuss such checks. Our goal here is only to provide a succinct survey, showing that the guidance offered by the principles of individuation is unclear, and limited at best. This last point is further confirmed in the following section, where we present our own objections to the redescription strategy – objections which are left entirely unaddressed by these principles – and then offer our own solutions.

To survey the principles of individuation, it is convenient to rely on Buchak's (2013, p. 123-136) threefold classification. One principle is contained in the following statement by Broome (1990, p. 103):

[13] Proponents of the strategy do not seem to have ever considered redescriptions based on *coarsenings* rather than refinements of the original outcomes.

"*Principle of Individuation by Justifiers* (PIJ). Outcomes should be distinguished as different if and only if they differ in a way that makes it rational to have a preference between them".[14]

Buchak extracts two other principles from less explicit formulations due to Dreier (1996) and Pettit (1991) respectively. In her words:

"*Principle of Individuation by Actual Preferences* (PIA). Outcomes should be distinguished as different if and only if the agent actually has a preference between them".

"*Principle of Individuation by Properties* (PIP). Outcomes should be distinguished as different if and only if they differ in regard to properties that are desired or undesired by the agent".

Starting with PIJ, we first observe that this principle is equivocal, since the expression "makes it rational to have a preference between (outcomes)" can mean two different things. This is because the rationality notion comes in two strata, first in terms of coherence conditions on the agent's preferences that take her desires and beliefs as *given*, and second in terms of higher-order properties that make these desires and beliefs *justifiable*. Sen's first example can easily illustrate this duality and how it impacts PIJ. If one adopts the first, *thin* notion of rationality, it is enough to point out with some plausibility that the agent applied a norm of politeness that overruled the formal choice conditions. One simply checks that the agent's choice of x from $\{x, y\}$ coheres with the choice of y from $\{x, y, z\}$ given the politeness assumption. Meanwhile, if one adopts the second, *thick* notion of rationality, one will have to inquire whether politeness was an appropriate attitude in the given situation (though in order to get a grip on this question, one will need to describe the latter more precisely). Even more systematically, one might have to inquire what social contexts make it appropriate to behave politely, and perhaps even what justifies politeness as a social institution. Clearly, a normative assessment of Sen's agent does not have to address the broader questions that the thick notion of rationality raises; and indeed, most of the work in decision theory is concerned with exploring the thin notion only.[15] Accordingly, property α must be evaluated in terms of this notion. The expression "makes it rational to have a preference between them", which oscillates between the two strata of rationality, should give way to a clumsier but better-

[14] In this statement and the ones that follow, "preference" has its ordinary language sense, i.e., it is synonymous with "strict preference" in decision theory.

[15] To use philosophical terminology that is currently widespread, this makes decision theory a brand of "Humeanism". The terminology adopted here of "thin" *versus* "thick" rationality comes from the methodology of the social sciences. It resembles the traditional philosophical distinction between "formal" and "substantial" rationality.

determined expression such as "makes it possible to postulate a preference between them and conclude that the choice made is rational, given that preference".

Dispelling the initial equivocation of PIJ does not salvage this principle from the objection that to rationalize an agent's choices from the outside can fail to capture what this agent's genuine reasons are. As Dreier (1996, p. 260) writes: "The fact that a set of preferences could be rationalized by attributing to the agent an unusual extra preference and individuating outcomes in an unusual way does not mean that those preferences are in fact rational. For the apparent violation might not be a result of any such unusual preference, so it might be a real violation". What Dreier has in view is a thin concept of rationality, and he is in effect arguing that this concept requires that we consider the agent who is actually presented in the counter-example, not an abstractly possible agent. It does not help to mention politeness as a possible reason for the violation of property α if this is not the reason moving *that* particular agent in *those* particular circumstances.[16]

This objection might justify turning to PIA; but if it is to guide actual normative discussions, this principle has its problems too. How is one to ascertain that the agent's actual preferences coincide with those hypothesized by the decision theorist? This question should in fact be subdivided as follows. How is one to ascertain the agent's ability to *distinguish* between outcomes as the theorist does? How is one to ascertain the agent's willingness to *rank* outcomes as the theorist does, granting that she accepts his distinctions? In principle, these are independent questions. For instance, the agent who chooses a piece of cake first could well recognize what politeness would instruct her to do, and nonetheless neglect this recommendation in her choices.

Existing defences of PIA do not usually separate these two levels of difficulties, and this may be because the first question is by itself difficult enough to answer. Implicitly focusing on this question, Dreier proposes asking the agent whether she individuates outcomes in the hypothesized way. This can work in simple cases like Sen's examples, but may be infeasible in other, more complex cases. It seems fruitless to ask the Allais agent whether she individuates the 0 outcome of Ball 11 differently in q^*_1 and q^*_2. The restatement of the agent's choice problem in Table 3 is a construct of the decision theorist, and it is far from obvious that the agent will adhere to it, let alone understand it. Admittedly, it is easier to

[16] See also Buchak (2013, p. 128-129). Bradley (2017, p. 23) dismisses PIJ, though for the different reason that the preferences involved in a principle of individuation can be normatively determined without being subject to a rationality constraint. In still another critical move, Bermudez (2009, p. 103-108) stresses that it is difficult to provide an independent grounding for the rationality concept involved in PIJ. Bermudez understands Broome as being after a thick rationality concept.

get from her the information that she did not choose q^*_1 because she would have regretted that choice if the unfavourable chance had been realized. However, this answer is not specific enough to warrant the conclusion that the agent and the theorist individuate outcomes identically. Anticipated regret can influence the agent's preferences in various ways, which might correspond to different modes of individuating outcomes, or not even warrant that regret or the lack thereof enters the definition of outcomes. Notice that experimentation is unlikely to do any better than Dreier's proposed questionnaire to establish empirically that the agent agrees with a particular individuation of outcomes.[17]

Let us now briefly turn attention to PIP. One wonders whether this principle is really autonomous. The "properties that are desired or undesired by the agent" constitute reasons why the agent has the preferences she has, and the decision theorist who means to apply PIA should obviously try to guess these reasons, and should in principle reach the same redescription whether he relies on PIA or PIP. Thus, in Sen's second example, if the decision theorist claims that the agent prefers "going and having tea when tea is the only offer" to "going and having tea when tea and cocaine are offered", this is because she believes that the undesired property of being shocked is what makes the difference between the two options in the agent's eyes.

All in all, if one had to make a choice among the principles listed by Buchak, it should be PIA. Although it runs into difficulties, it is not open to such a stringent objection as that which Dreier rightfully raises against PIJ. Moreover, it delineates a possible role for this principle: for, although not defensible independently, PIJ can enter stage when it is impractical or dispensable to collect empirical data on how the agent would view the redescription. In this case, following a standard heuristic of charitable assignment, the decision theorist can infer the agent's attitude from the mere fact that it rationalizes her observed choices, in the thin rationality sense. To return to Sen's second example, it is clear enough from the comparison of opportunity sets what reason the agent has to revise her choice of accepting the invitation, and at least in such easy cases, PIJ can do the job of justifying the redescription without any recourse to the empirical investigation PIA normally requires.[18]

[17] Buchak (2013, p. 131-132) also discusses this problem of finding empirical evidence, though she chooses to do that in connection with PIP rather than PIA.

[18] That PIJ can play this auxiliary role seems to be the grain of truth in Broome's own position. One may interpret him as reacting to the difficulty of finding empirical evidence that would make PIA operative. He does, however, overreact by promoting PIJ as if it were a self-sufficient principle.

5. Where the main objections lie and the need for a theoretical expansion

The principles of individuation exhaust the received answer to the worrying objection that the redescription-of-outcomes strategy is too easy to implement. Irrespective of their comparative merits, these principles are objectionable on the ground that they take the proposed redescriptions *just as they stand*, as if these redescriptions had received their proper formulation. But this cannot be the case; the representative examples of sections 2 and 3 are at best heuristic, at worst anecdotal. It is not simply that they are too easy to come by. More importantly, they raise the following objections.

Redescriptions of outcomes can be incomplete. While Sen's examples seem to eschew this objection, the Allais paradox is clearly open to it. In Table 3, the redescription in terms of regret is carried out on only two outcomes, i.e., those of p^*_2 and q^*_2 on Ball 11, though consistently it should be made also for the outcomes of p^*_1 and q^*_1 on Balls 1-10. For if any of these ten balls is drawn, a regretful agent should get annoyed about having chosen p^*_1 (p^*_2) rather than q^*_1 (resp.q^*_2). When Table 3 is completed in this way, the final conclusion depends on how the balance is struck between the regret components of the different outcomes.

Redescriptions of outcomes salvage the axioms or principles of decision theory in a logically trivial way. Both earlier applications elude the counter-examples by having the threatened decision-theoretic principles hold vacuously. Thus implemented, the strategy cannot score great victories. One would rather have the principles *imply* the counter-examples after a more thorough reformulation. If, in each case, the principle targeted by the counterexample is the key norm of rationality in the choice context of interest; if, furthermore, the presented counterexample offers a key insight into what rational choice amounts to in this context; and if, finally, there is to be any sort of harmony between these two tenets that are built into the normative discussion – then the goal of the strategy should be to have the principle not just *allow for* (which tolerates the principle holding vacuously), but *demand* (which excludes this case) the paradoxical preferences, once suitably redescribed.[19] In this way, the counter-examples would be turned into confirmatory examples instead of being

[19] This requirement should not be interpreted more strongly than intended. Each paradox may be decomposed in antecedent and consequent preferences. What we envisage here is that *given the antecedent preferences*, the relevant principle of rationality imply the consequent preferences. That is naturally not to say that rationality should mandate, rather than merely allow for, the antecedent preferences themselves.

simply put aside. The classic deductive account of explanation, for instance, does not permit the *explanans* to hold vacuously; only a proper deduction of the *explanandum* can warrant the claim that an explanation has taken place. Curiously, even those who have rightly stressed that redescription is too easy a move do not seem to have considered the logical sense of triviality raised here. The triviality problem they attempt to address by resorting to the "principles of individuation" is of an entirely different nature – it relates to the large number of considerations that could be brought to bear upon the definition of an outcome and the difficulty in checking which ones are relevant.

Redescriptions are theoretically insular. Redescriptions are typically geared at specific counter-examples to specific decision-theoretic statements. But in decision theory, counter-examples to a given condition often come together in a group, and it is not unusual that one and the same counter-example should hit more than one condition. Section 2 illustrates these two points: as Sen has made clear, his counterexamples are mutually supportive and they target the standard theory of choice functions over and beyond the binariness condition. Moreover, redescriptions claim to uncover features that have often been explored elsewhere in decision theory, albeit in a different way, and the relationship with these theoretical developments must be discussed. The full-fledged theories of regret or disappointment avoidance available in up-to-date decision theory are cases in point. This problem of missing theoretical links appears the most important in our list, because if it can be resolved, solutions to the incompleteness and the triviality problems might also be forthcoming. Thus we recommend a single general recipe, which is to judge each case on the merits of its *theoretical expansion.* We close this section by programmatically indicating (before actually detailing in the next two sections) how this recipe should be applied to the Sen and the Allais examples.

Over and beyond the anecdotal suggestion that the Sen agent is polite in one case, and scared in another, her violations of property α appear to be due to the phenomenon of *menu-dependence*, as introduced in section 2. The question that a theoretical expansion could help to resolve here, then, is whether this concept can be properly handled by a suitably revised theory of choice functions, such as one of the currently available theories of menu-dependent choice that have been developed for this purpose. If this turned out to be the case, the redescription would have to be praised for its fruitfulness; if it failed, one would have to discard it from consideration.

While the Allais agent is said to fear regret or disappointment, decision theory has axiomatically characterized preferences under uncertainty that are guided by the avoidance of these feelings. An obvious line of theoretical expansion here is to check what becomes of the

redescription when it is restated in terms of these constructions. The question here is whether such an approach can turn the allusive reference to negative feelings into a proper explanation of the Allais paradox, while simultaneously preserving the STP.

6. Theoretically expanding on redescriptions in Sen's case

The counter-examples of section 2 have given rise to a handful of theoretical studies, all of which have some relevance to the topic of this paper.[20] Concerning the first, politeness counter-example, we select a paper by Bossert and Suzumura (2009), which develops a formalism of choice functions for the case when an external rule of conduct, or "norm" is at work. They capture the role of the norm by defining a function f that associates to each subset of options S those options in S which the norm *forbids*. Thus, for each subset of options S, S-$f(S)$ is the *permissible* subset of S. By assumption, the choice function C obeys the norm, i.e., for all subsets of options S, $C(S)$ is included in S-$f(S)$. Bossert and Suzumura show that the standard theory of choice functions applies to C with only limited modifications, all of which connect with the presence of the constraining norm. They do not consider property α; but it can be reformulated to fit their framework. It then says that if the permissible subset of S is included in the permissible subset of S', y is chosen from S', and y is a permissible element of S, then y is chosen from S. The usual condition simply says that if S is included in S' and y is chosen from S', then y is chosen from S.

Bossert and Suzumura's account of norms in terms of an additional function f may seem oversimplified and their extension of the standard theory, not to come as a great surprise; but they can be credited with making the redescription in Sen's first example part of a proper decision-theoretic framework. This extension responds to the insularity problem, and since the redescription cannot be accused of being incomplete, it remains only to check the logical triviality problem. Unfortunately, the reformulated property α is only vacuously satisfied. If the two choices $C(\{x, y\})=\{x\}$ and $C(\{x, y, z\})=\{y\}$ satisfy it, this is because y is *not* permissible in $\{x, y\}$, so that the antecedent clause of property α does not hold.[21]

[20] Bhattacharyya, Pattanaik, and Xu (2011) survey the main contributions.

[21] Baigent and Gaertner (1996) formalize the politeness explanation of the first example differently from Bossert and Suzumura. They interpret the agent's choices as being driven by the rule "never choose the uniquely largest", and axiomatize compliance with this rule in terms of properties of choice functions. The relevant properties are interestingly close to the standard consistency conditions. Heuristically, this work accords with the redescription that takes the two choice sets to be {next-to-largest piece, largest piece} and {smallest piece, next-to-largest piece, largest}. This treatment is

Concerning Sen's second, tea and cocaine example, we rely on a paper by Bossert (2001), which for the most part is a theoretical development of this very case. The guiding idea is that informational menu-dependence is best handled by moving from the theory of decision under certainty to that of decision under uncertainty. The actual choice is between "going and having tea" (x), "not going" (y), and "going and having cocaine" (z). However, the argument goes, this statement of *options* does not fully inform one of which *outcomes* they lead to. These should be stated as "going and having tea at a place where no cocaine is consumed" (a), "going and having tea at a place where cocaine is consumed" (b), "not going" (c), and "going and having cocaine" (d). While y is identical with c, and z with d, x has a more complex association on the outcome side, since it can result in either a or b; this is where the uncertainty model comes in. The discrepancy observed for x suggests a general treatment, which is to redefine all options as *sets of outcomes*, i.e., x will be replaced by $\{a, b\}$, y by $\{c\}$, and z by $\{d\}$. Bossert proceeds to develop a theory of choice functions for such set-valued options. Except for his more abstract formulation, the consistency conditions he defines are identical to the standard ones. For the example which concerns us here, his theory makes it possible to explain the agent's choices in terms of a choice function that can be rationalized by an ordering. Again, this explanation hinges on the assumption that the choices among the original options can be mapped onto choices made among the set-valued options.

While remaining faithful to the heuristics of redescribing outcomes rather than other components of the choice problem, this construction does a good job at integrating an initially isolated example into a framework of decision theory. Thanks to the ingenious step from the theory of certainty to that of uncertainty, it provides a way of handling many cases of informational menu-dependence. Thus the insularity problem is elegantly solved, and since again there can be no accusations of incompleteness, the only pending problem is logical triviality. Regrettably, however, Bossert's construction does not in general surmount this problem. To see this, let us spell out the ordering of sets of consequences that rationalizes the choices made by Sen's agent. These choices fix the ranking of singletons, i.e., $\{a\}\mathrm{P}\{c\}\mathrm{P}\{b\}\mathrm{P}\{d\}$, where P represents strict preference. Let us add the assumption that $\{a, b\}\mathrm{P}\{c\}$, i.e., the agent would rather go to a place which may or may not be commendable than not going. This assumption is not implausible in itself, and it is in any case perfectly compatible with the motivations usually ascribed to Sen's agent. Reinterpreting the initial choices in terms of the partial ordering thus specified leads to the choice of $\{a, b\}$ rather than

unfortunately only of limited interest to our enquiry because it is restricted to a special class of choice problems and does not illuminate the effect of following a norm generally.

$\{c\}$, and then to the choice of $\{c\}$ rather than $\{b\}$ and $\{d\}$. Neither $\{a, b\}$ nor $\{a\}$ are part of the menu of the second choice; this is because of the informational effect of adding $\{d\}$. Thus, the menu of the first choice is not included in the menu of the second choice, and once again, property α holds only vacuously.

7. Theoretically expanding on redescriptions in Allais's case

Let us take up the familiar response to the Allais paradox in terms of *regret*. When the agent chooses q^*_1 and Ball 11 is drawn, she is said to feel regret. This claim concerns her *ex post* situation, and it matters to the assessment of the sure-thing principle only if it is part of an account of her *ex ante* deliberation. So the next claim is that the Allais agent chooses p^*_1 because she anticipates she could feel regret with q^*_1. But this will remain a promissory note as long as it is not articulated in terms of a well-defined criterion of regret-avoidance. This criterion would have to take account of all the boxes in Table 3 that contain a regret element. The redescription strategy here is both insular and incomplete.

Let us now see how, for example, the *minimax*-regret criterion would operate.[22] Informally speaking, this criterion says that the agent makes a twofold computation. First, she computes all possible regrets for each prospect taken separately. She does that by considering each state in turn and taking the difference between two quantities: on the one hand, the *best* possible value of the outcome she could get if she let the prospects vary while keeping the state fixed, and on the other hand, the value of the outcome associated with the given prospect and the given state. Second, the agent selects the *maximum* of the regrets so computed to represent the negative value of the prospect itself, and finally selects one of the prospects having *minimum* negative value.

This criterion formalizes the regret notion in what seems to be the appropriate comparative way, and its final reliance on a minimax comparison agrees with the intuition underlying the paradox that one must be "very prudent" (Allais's words) to prefer p^*_1 to q^*_1. However, the criterion does not deliver the desired result, as the computation stated below readily shows. We give two tables of regrets, one for the pair (p^*_1, q^*_1), and the other for the pair (p^*_2, q^*_2); they include a utility function u indicating that the agent compares the *values*

[22] Savage (1951) connected *minimax* regret with statistical theory. We only need the elementary version of Milnor (1954).

of monetary outcomes, not these outcomes themselves. If this function exhibits diminishing marginal utility at a sufficiently quick pace, then

$$u(500) - u(100) < u(100) - u(0),$$

and Table 4.1 secures the desired conclusion that the agent prefers p^*_1 to q^*_1. But by the same token, Table 4.2 – being identical to Table 4.1 – delivers the unwelcome conclusion that the agent prefers p^*_2 to q^*_2. (In these tables, we have underlined the maximum value of regrets for each prospect.) Hence, the *minimax*-regret criterion does not make it possible to reproduce the Allais paradox. On reflection this is unsurprising, because this criterion, relying on state-by-state comparisons, does satisfy the sure-thing principle. What emerges from its application is one of the two EU-conforming solutions to the two choice problems.

Table 4.1	Balls 1-10	Ball 11	Balls 12-100
p^*_1	<u>u(500)-u(100)</u>	0	0
q^*_1	0	<u>u(100)-u(0)</u>	0
Table 4.2			
p^*_2	<u>u(500)-u(100)</u>	0	0
q^*_2	0	<u>u(100)-u(0)</u>	0

Table 4

More sophisticated criteria of regret avoidance do not reproduce the Allais paradox either.[23] As all of these criteria rely on state-by-state comparisons, they compute regrets in the same way for p^*_2 and q^*_2, and for p^*_1 and q^*_1. In particular, there is no way of capturing the contrast of "0 with regret" and "0 without regret" in Table 3. There is a gap between the intuitive ground for the redescription, and what decision theory achieves when it is properly applied.

Instead of taking regret as a basis for description, one could with equal psychological plausibility select *disappointment*. Both feelings are negative and arise *ex post* from comparing the actual outcome with other outcomes that could have obtained; but the

[23] Typically, these criteria replace *minimax* by expected value at the final stage of the deliberation, and include an evaluation of the outcome itself besides that of the regret attached to the outcome; see, e.g., Bell (1982). They do not help recover the Allais paradox because they satisfy the sure-thing principle. If Loomes and Sugden's (1982) regret theory does not satisfy this principle, and can indeed recover the Allais paradox, this is because they make the special assumption of *independent variation*, according to which the prospects being compared give rise to stochastically independent drawings.

counterfactual reference is not the same in the two cases. For a given prospect, *regret* effects a comparison in any given state with the outcome that another prospect would have brought about in this state, whereas for a given prospect, *disappointment* effects a comparison in any given state with the outcomes that other states would have brought about for this prospect. If one is content with redescribing the outcomes of Ball 11 as "0 with disappointment" for q^*_1, and "0 without disappointment" for q^*_2, the twofold objection of incompleteness and insularity arises again.

As a possible remedy, let us try a criterion for disappointment avoidance that is due to Bell (1988).[24] In a simplified version that is sufficient for our purposes, this criterion relies on a disappointment function $d(.)$ and defines the utility of the monetary outcome $x(p, s)$ of a lottery p in a state s as the quantity $u(x(p, s)) - d(E_p - x(p, s))$, where E_p stands for the expected monetary value of p. Tables 5.1 and 5.2 state the disappointment values of the four prospects accordingly; expected monetary values are computed using the probability values of the Allais lotteries.

Table 5.1	Balls 1-10	Ball 11	Balls 12-100
p^*_1	$\underline{u(100) - d(0)}$	$\underline{u(100) - d(0)}$	$\underline{u(100) - d(0)}$
q^*_1	$u(500)$	$\underline{u(0) - d(139)}$	$u(100) - d(39)$
Table 5.2			
p^*_2	$u(100)$	$u(100)$	$\underline{u(0) - d(11)}$
q^*_2	$u(500)$	$\underline{u(0) - d(50)}$	$u(0) - d(50)$

Table 5

Given that u and d have to be increasing functions, it is unproblematic to find the minimum value associated with each prospect, and, after applying the same "very prudent" reasoning as we did earlier (represented this time by a *maximin*, not a *minimax* computation), to conclude that the agent will choose p^*_1 and q^*_2. Thus, with a modicum of disappointment theory, it is easy to reproduce the Allais paradox. The inequality $u(0) - d(139) < u(0) - d(50)$ provides a graded version of the initial contrast between "0 with disappointment" and "0

[24] There exist other such criteria; see Loomes and Sugden (1986) and Gul (1991). They would lead qualitatively to the same conclusion here.

without disappointment". However, disappointment theory involves comparisons made across states, not state by state as in regret theory, and so plays havoc with the STP. In the end the two theories fail in different ways, but for the same reason, viz. neither can *both* account for the paradox and salvage the principle. Compare this situation with that reached concerning property α. There, the theoretical principle was at least trivially compatible with its counterexamples; but here even this is not possible. The failure is clearly worse in the Allais case than in the Sen case.[25] A general lesson from the foregoing is that taking due account of the need for a theoretical expansion, one should not assume that, for every normative puzzle, there exists a redescription strategy to dissolve it away – even in a logically trivial fashion.

8. Redescriptions in Ellsberg's case

At this stage, one might wonder whether any any redescription could pass our proposed test. To provide more evidence on this question, we turn to the Ellsberg paradox and how the redescription strategy has been used to overcome it. As will be shown, in this case the strategy passes the test of theoretical expansion, i.e., it connects well with some of the decision-theoretic work that has been pursued on the paradox. However, there are significant differences between the initial description sketches and the properly expanded analytical proposal. What is more – and this should further moderate the praise – the strategy here is not limited to redefining outcomes; it encompasses the choice problem as a whole, which turns out to be the decisive move.

Ellsberg's (1961) paradox is concerned with subjective expected utility (SEU) theory of the Savage brand. It challenges the EU representation that this theory derives for preferences on its characteristic objects, i.e., uncertain prospects. More specifically, it can be shown to contradict both the STP and the property of *probabilistic sophistication*, which is implied by Savage's SEU representation, i.e., the property that a subjective probability measure enters the utility representation of preferences.[26]

[25] This seems to be well enough to dispose of the redescription-of-outcomes strategy in the Allais case. But we could have followed another direction of theoretical expansion by investigating whether other counter-examples proposed by Allais and followers could similarly suggest an explanation in terms of negative feelings. The short answer is that this is not always the case; see the established list of counter-examples in Larsson and MacCrimmon (1979), Machina (1987), and (with a philosophy of science perspective) Mongin (2009).

[26] Probabilistic sophistication has become a focus of attention in decision theory after Machina and Schmeidler (1992) axiomatized in a more general form than the special form it takes in Savage's SEU representation theorem.

In a simplified variant, the paradox goes as follows. A three-ball urn is known to contain one red ball and two other balls that are either black or yellow in unknown proportions. One ball is to be drawn from the urn and the agent is asked to make two successive choices, each of which involves two bets on the colour of the ball. The first choice is between the bets that the colour is red (f_1) and that it is black (g_1), and the second choice is between the bets that it is red or yellow (f_2) and that it is black or yellow (g_2). The two bets are to be settled at the same time, simply by observing the colour of the ball to be drawn. Ellsberg's own solution, for which he also made some casual empirical verifications, is (f_1, g_2). The following argument makes it clear that this pair of choices clashes with probabilistic sophistication. The standard connection between monetary bets and subjective uncertainty is that, when faced with two bets, each of which puts a given stake on a different event, the agent prefers the bet on the event she regards as being more likely.[27] By this rule of translation, the Ellsberg agent takes "red" to be more likely than "black" (first choice); however, she also takes "black or yellow" to be more likely than "red or black" (second choice), and this contradicts the additivity of probability.

That Ellsberg preferences violate the STP can be seen directly from the next table, in which bets are listed as rows and states of the world as columns. For concreteness, the stake in each bet is fixed at $100.

	Red	Black	Yellow
f_1	100	0	0
g_1	0	100	0
f_2	100	0	100
g_2	0	100	100

Table 6

Since f_2 and g_2 can be obtained from f_1 and g_1 simply by changing their common outcome on the state Yellow from 0 to 100, preferring f_1 to g_1 and g_2 to f_2 violates the STP.

Ellsberg gave a psychological explanation for these preferences, which he also meant to be an argument for their reasonableness. As this explanation goes, the agent prefers betting on events that can be given precise probability values, such as Red in the first choice and

[27] We state this connection informally, but it is actually implied, as an intermediary result, by Savage's postulates.

Black or Yellow in the second, to betting on events that have imprecise probability values, such as Black in the first choice and Red or Yellow in the second. Exactly 1/3 of the balls are red, exactly 2/3 of the balls are black or yellow, but the proportion of black balls is 0, 1/3, or 2/3, and that of red or yellow balls is 1/3, 2/3, or 1. The ensuing literature has explored this account in terms of *ambiguity aversion* as thoroughly as it has explored the certainty and complementarity effects in the case of the Allais paradox.[28] However, the redescription literature (which is incidentally much less abundant here than it is on either Sen or Allais) points in a different direction, which we now explain.

As a first preliminary step, one may observe that the agent's uncertainty is incompletely described, and therefore propose as a remedy to refine the states of the world. This can be done by specifying not only the colour of the ball drawn, as in Table 6, but also from which particular set of coloured balls it is drawn. In Table 7 below, the states of the world are redefined in terms of both pieces of information, e.g., R&RBB means that Red is the colour of the ball drawn and that the composition of the urn is one red and two black balls. The latter information only concerns the number of coloured balls in the urn, irrespective of the order in which colours may be listed: i.e., instead of RBY, we could have written RYB, YRB, and so on.

	R&RBB	R&RYY	R&RBY	B&RBB	B&RBY	Y&RYY	Y&RBY
f_1	100	100	100	0	0	0	0
g_1	0	0	0	100	100	0	0
f_2	100	100	100	0	0	100	100
g_2	0	0	0	100	100	100	100

Table 7

This redescription cannot dissolve the conflicts with the STP and probabilistic sophistication since it simply consists in subdividing the initial states Red, Black, and Yellow. But it can subsequently be combined with the more usual step of refining outcomes. For concreteness, we will first follow an assumption made in most attempts at reformulating Ellsberg's paradox by complexifying states of the world: this is the symmetry assumption that for

[28] See Machina and Siniscalchi (2014) for a review of the ambiguity literature.

every composition of the urn, each ball has an equal chance of being drawn. Thus, one can say exactly why the chances of winning $100 are not the same in, say, a bet on B&RBB and a bet on B&RBY – they are 2/3 in the former and 1/3 in the latter. Consistently with this assumption, a natural refinement of outcomes consists in associating each money amount with the chance of drawing a ball of a given colour, from an urn of a given composition.[29] As a second preliminary step, the next table implements this refinement of outcomes.

	R&RBB	R&RYY	R&RBY	B&RBB	B&RBY	Y&RYY	Y&RBY
f_1	(100; 1/3)	(100; 1/3)	(100; 1/3)	(0; 2/3)	(0; 1/3)	(0; 2/3)	(0; 1/3)
g_1	(0; 1/3)	(0; 1/3)	(0; 1/3)	(100; 2/3)	(100; 1/3)	(0; 2/3)	(0; 1/3)
f_2	(100; 1/3)	(100; 1/3)	(100; 1/3)	(0; 2/3)	(0; 1/3)	(100; 2/3)	(100; 1/3)
g_2	(0; 1/3)	(0; 1/3)	(0; 1/3)	(100; 2/3)	(100; 1/3)	(100; 2/3)	(100; 1/3)

Table 8

Clearly, Table 8 still displays a conflict between the Ellsberg preferences and the STP; but as we now explain, it can serve as a basis for solving that conflict. The two components of uncertainty taking centre stage in Table 8 – i.e., that which is relative to the colour composition of the urn, and that which is relative to which ball in the urn is drawn – could be separated even more clearly. A simple but powerful way to do so is to assume that each ball has not only a colour, but also a painted serial number, viz. 1, 2, or 3, with 1 being painted only on a red ball, and 2 and 3 being painted sometimes on a black ball, sometimes on a yellow one. We will also assume that the colour composition of the urn specifies a colour for each numbered ball (thus giving more than just frequency information concerning the colours, as was previously the case). This will be expressed by following the serial order when the composition is stated; e.g., RYB will now indicate that the yellow ball has number 2 and the black ball has number 3, while RBY interchanges these numbers. The device of serially ordering the balls to fully separate the two components of uncertainty in the Ellsberg paradox is due to Ergin and Gul (2009) and taken up by Machina (2011). It makes it possible to have an "orthogonal" representation – Machina's term – of two forms of uncertainty taken together.[30] The next tables state this representation for the four bets in succession.

[29] A related but different approach has recently been explored by Bradley (2016, p. 234-238, and, in less detail, 2017, p. 175-177).
[30] In more mathematical terms, a Cartesian product representation.

3	0	0	0	0
2	0	0	0	0
1	100	100	100	100
	RBB	RBY	RYB	RYY

Table 9.1: f_1

3	100	0	100	0
2	100	100	0	0
1	0	0	0	0
	RBB	RBY	RYB	RYY

Table 9.2: g_1

3	0	100	0	100
2	0	0	100	100
1	100	100	100	100
	RBB	RBY	RYB	RYY

Table 9.3: f_2

3	100	100	100	100
2	100	100	100	100
1	0	0	0	0
	RBB	RBY	RYB	RYY

Table 9.4: g_2

These "orthogonal" representations may suggest more than one direction of analysis. Machina (2011) reaches interesting conclusions by taking balls 1, 2, and 3 to be equiprobable, thus reproducing in the new framework the symmetry assumption that was used in our preliminary Table 8. As the Ellsberg paradox persists even in the absence of this assumption, we now prefer to be more general and take the serial number of the ball to be uncertain in the same sense as the colour composition of the urn; that is, we treat the two components of uncertainty as being of the same nature instead of assuming that one takes a

subjective form (with unknown probabilities) and the other an objective form (with known probabilities). This being granted, the "orthogonal" representations permit one to define three different kinds of uncertain prospects, depending on which mappings from states of the world to outcomes one considers. A mapping according to *Definition 1* takes states of the world, as provided by the full "orthogonal" representation, i.e., (RBB, 1), (RBB, 2), (RBB, 3), (RBY, 1), … , to money outcomes. A mapping according to *Definition 2* takes states of the world as provided by the columns of the "orthogonal" representation – i.e., the compositions RBB, RBY, RYB, RYY – to uncertain prospects of a relevant kind viewed as outcomes; these prospects are mappings from the serial numbers to money outcomes. Finally, a mapping according to *Definition 3* takes states of the world as provided by the lines of the "orthogonal" representations – i.e., the serial numbers 1, 2, 3 – to uncertain prospects of a relevant kind viewed as outcomes; the prospects in question are mappings from the compositions to money outcomes. Now, for each of the three definitions, one should ask whether the STP principle holds. It is tedious, but straightforward to check that the principle is violated with definition 1, that is the most encompassing of the three and the only one to correspond to the uncertain prospects usually mentioned in Ellsberg's paradox. Following either definition 2 or definition 3, however, the principle vacuously holds, and the argument against probabilistic sophistication is blocked.

At this juncture, one can use advanced decision-theoretic tools to convert this already striking insight into a non-trivial resolution of the conflict between the STP and the Ellsberg preferences. The theories of second-order expected utility developed by Klibanoff, Marinacci, and Mukerji (2005), Nau (2006), and Ergin and Gul (2009) implicitly rely on the orthogonal analysis just introduced, and they push it further as we detail next. Their representation theorems involve *an SEU representation for each component of uncertainty separately, but a non-EU representation for the "orthogonal" uncertainty as a whole*. The latter representation takes the form of a (second-order) EU function that varies with the values of a non-linear function, which is defined on the values of a (first-order) EU function. This two-stage formula does not collapse into a single EU formula because of the non-linearity of the function that intervenes between the two EU levels. The formula is hierarchical, i.e., it takes one component of uncertainty to be primary, and the other to be secondary, in the sense that it is an uncertainty about which uncertainty prevails at the primary level. This asymmetry is an extraneous feature, compared with the previous analysis. However, for relevant parameter values, Ellsberg's pair of choices can be deduced from the second-order expected utility formula. Last but not least, as the very structure of the model indicates, this is obtained through a rather conservative departure from SEU. One

can interpret this as fulfilling the protective intentions initially motivating the redescription strategy.[31]

Thus, it would seem that, in the particular case of the Ellsberg paradox, the redescription strategy can overcome its usual defects. A proper deduction takes place, the redescription is carried out completely, and it is connected with the rest of decision theory. However, one must realize, and indeed stress, that this by and large satisfactory conclusion depends on taking the strategy much more broadly than is usually the case. The impetus for the analysis comes from the realization that the agent's uncertainty is not fully described, and in the final stage of this analysis the descriptive refinements are concentrated on the definitions of states and prospects, whereas the definition of outcomes is left unchanged.

9. Conclusions

The basic axioms or formal conditions of decision theory, especially the ordering and related conditions such as property α, and the axioms that decompose the EU formula in its various forms, face a number of counter-examples, or "paradoxes" as the common expression goes. The strategy of redescribing outcomes has been devised to deflect the subclass of counter-examples that appear to be normatively justifiable, thus to claim that there is in fact no paradox, i.e., no tension between mutually incompatible normative intuitions. Philosophers lay special emphasis on this strategy, which has attracted less attention from decision theorists themselves. Motivated by the obvious worry that it may be too easy to implement and may thus deprive the statements of their content, philosophers have accordingly devised safeguards in the form of "principles of individuation". In the present paper we have developed a more radical critique of the redescription-of-outcomes strategy, observing from a sample of applications that the strategy has three major defects, to wit logical triviality, incompleteness (in at least some cases), and theoretical insularity. Going beyond what the "principles of individuation" permit, we have proposed to accept only those redescriptions which manage to overcome these defects once they are developed theoretically. To implement this criterion entails checking each case on its respective merits, which delivers a nuanced assessment of which redescriptions are acceptable and which

[31] In the theories of second-order expected utility, the STP holds *non-trivially* at each level of uncertainty. This is necessary for these theories to obtain a full-fledged SEU representation at each level. This is one of the senses in which these theories push the analysis of "orthogonal" representations further than the analysis made in the previous paragraph.

should be refused. We have shown that the redescription strategy was only partly satisfactory in the case of Sen's counterexamples to property α and the related binariness property, and that it failed in the case of Allais's paradox for EU theory. The redescription proposed in the case of Ellsberg's paradox against EU theory comes the closest to satisfying our requisite.

Although this is a limited sample, it is a representative one, and some general conclusions emerge. For one thing, the logical triviality objection may be the most difficult defect to overcome; it mars otherwise serious attempts at theoretically expanding on redescriptions. For another thing, the status of a redescription improves – even dramatically, as testified to by the Ellsberg application – when it is extended to more definitional components of the choice problems than just the outcomes – such as the states of the world and the prospects in the Ellsberg case. Last but not least, to overcome its natural defects, the revision endeavour should extend to the very theoretical statement that the redescription is meant to protect; it is dubious that it can have any explanatory value when this is not done. In the case of counter-examples to EU theories, the strategy, when properly developed, will almost inevitably lead one to borrow from the rich collection of non-EU theories. If the deviation from the EU theories is kept to a minimum, this will still fulfil the protective intentions initially motivating the strategy. Our treatment of Ellsberg's paradox is again to the point here, as the second-order expected utility theories it resorts to are economical departures from standard EU theories. One way or another, then, the overall lesson of the foregoing is that there is no easy way out of the paradoxes of decision theory. The dilemmas which they present to the work on practical rationality cannot be escaped, and should be faced squarely.

To close with a broader perspective on the issues of this paper, it may be added that our requirement for a proper theoretical expansion, together with the main lesson just highlighted, are consistent with and indeed were in part suggested by elementary considerations drawn from the philosophy of science. The objections raised against the redescription strategy are not unlike those raised against the dismissal of empirical refutations in science. This dismissal can take the form of claiming that the observed conditions did not match the antecedent clause of the conditional law submitted to the test. Although this claim may well be empirically supported, it discards the test result without explaining it – an objection that connects with our complaint of logical triviality. The dismissal can also take the form of claiming that some parameter values were not those required for the test to be informative – which indirectly connects with our complaints that

redescriptions can be incomplete and are theoretically insular. Briefly put, responses to either empirical refutations or normative counter-examples are in danger of being *ad hoc* whenever they are isolated from the existing theories in the field. The final judgement on their appropriateness concerns whether they still meet their original protective purpose once they are developed in terms of these theories. In the end, to draw on another basic analogy between the empirical and the normative, we may say that in normative decision theory like elsewhere, conflict between theory and recalcitrant data can never be seriously avoided for long.

References

Allais, M. (1953), "Le comportement de l'homme rationnel devant le risque", *Econometrica*, 21, p. 503–46.

Allais M. (1979), "The So-Called Allais Paradox and Rational Decisions under Uncertainty", in M. Allais and O. Hagen (eds), *Expected Utility and the Allais Paradox*, Dordrecht: D. Reidel, p. 437–683.

Baigent, N. and W. Gaertner (1996), "Never Choose the Uniquely Largest: A Characterization", *Economic Theory*, 8, p. 239-249.

Bhattacharyya, A., P.K. Pattanaik, and Y. Xu (2011), "Choice, Internal Consistency and Rationality", *Economics and Philosophy*, 27, p. 123-149.

Bell, D.E. (1982), "Regret in Decision Making under Uncertainty", *Operations Research*, 30, p. 961-981.

Bell, D.E. (1985), "Disappointment in Decision Making under Uncertainty", *Operations Research*, 33, p. 1-27.

Bermudez, J.L. (2007), *Decision Theory and Rationality*, Oxford, Oxford University Press.

Bossert, W. (2001), "Choices, Consequences, and Rationality", *Synthese*, 129, p. 343-369.

Bradley, R. (2016): “Ellsberg's Paradox and the Value of Chances”, *Economics and Philosophy*, 32(2), p. 231–248.

Bradley, R. (2017), *Decision Theory with a Human Face*, Cambridge, Cambridge University Press.

Bradley, R. and O. Stefansson (2017), «Counterfactual Desirability», *British Journal for the Philosophy of Science*, 68, p. 485-533.

Broome, J. (1991), *Weighing Goods*, Oxford, Blackwell.

Buchak, L. (2013), *Rationality and Risk*, Oxford, Oxford University Press.

Dhongde, S. and P.K. Pattanaik (2010), "Preference, Choice, and Rationality", in C. Morris (ed.), *Amartya Sen*, Cambridge, Cambridge University Press. (A volume in the series, Contemporary Philosophy in Focus).

Dreier J. (1996), "Rational Preference: Decision Theory as a Theory of Practical Rationality, *Theory and Decision*, 40, p. 249–276.

Ellsberg, D. (1954), "Classic and Current Notions of "Measurable" Utility", *Economic Journal*, 64, p. 528-556.

Ergin, H. and F. Gul (2009), "A Theory of Subjective Compound Lotteries", *Journal of Economic Theory*, 144, p. 899-929.

Gul, F. (1991), "A Theory of Disappointment Aversion", *Econometrica*, 59, p. 667-86.

Hammond, P.J. (1988), "Consequentialist Foundations for Expected Utility", *Theory and Decision*, p. 25-78.

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

Loomes, G. and R. Sugden (1982), "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty", *Economic Journal*, 92, p. 805-824.

Loomes, G. and R. Sugden (1986), "Disappointment and Dynamic Consistency in Choice under Uncertainty", *The Review of Economic Studies* 53, p. 271–282.

Luce, R.D. and Raiffa, H. (1957), *Games and Decisions*, New York, Wiley.

MacCrimmon K.R. and S. Larsson (1979), "Utility Theory: Axioms versus 'Paradoxes'", in M. Allais and O. Hagen (eds), *Expected Utility and the Allais Paradox*, Dordrecht, D. Reidel, p. 333–409.

Machina M.J. (1987), "Choices Under Uncertainty: Problems Solved and Unsolved", *Journal of Economic Perspectives*, 1, p. 121–154.

Machina M.J. (1989), "Dynamic Consistency and Non-Expected Utility Models of Choice under Uncertainty", *Journal of Economic Literature*, 27, p. 1622–1668.

Machina, M.J. (2011), "Event-Separability in the Ellsberg Urn", *Economic Theory*, 48, p. 425-436.

Machina, M.J. and D. Schmeidler (1992), "A More Robust Definition of Subjective Probability", *Econometrica*, 60, p. 745-780.

Machina, M.J. and M. Sinischalchi (2014), "Ambiguity and Ambiguity Aversion", *Handbook of the Economics of Risk and Uncertainty*, 1, ch. 13, p. 729-807.

Milnor, J. (1954), "Games against Nature", ch. IV, in R.M. Thrall, C.H. Coombs, R.L. Davis (eds), *Decision Processes*, Wiley, New York, 1954, p. 49-59.

Mongin, P. (2009), "Duhemian Themes in Expected Utility Theory", in A. Brenner and J. Gayon (eds), *French Studies in the Philosophy of Science,* New York, Springer, 2009, ch. 13, p. 303-357.

Mongin, P. (2019), "The Allais Paradox: What it Became, What it Really Was, What it Now Suggests to Us", *Economics and Philosophy*.

Nau, R. (2006), "Uncertainty Aversion with Second-Order Utilities and Probabilities", *Management Science*, p. 136-145**.**

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton, Princeton University Press (2nd revised ed., 1947).

Neumann, M. (2007), "Choosing and Describing: Sen and the Irrelevance of Independent Alternatives", *Theory and Decision*, 63,79–94.

Pettit, P. (1991), "Decision Theory and Folk Psychology", ch. 4 in M. Bacharach and S. Hurley (eds), *Foundations of Decision Theory*, Oxford, Blackwell, p. 147-173.

Raiffa, H. (1968), *Decision Analysis*, Reading, Mass., Addison-Wesley.

Savage, L.J. (1951), "The Theory of Statistical Decision", *Journal of the American Statistical Association*, 46, p. 55–67.

Savage, L.J. (1954), *The Foundations of Statistics*, New York, Dover (2nd edn, 1972).

Sen, A.K. (1982), *Choice, Welfare and Measurement*, Oxford, Blackwell.

Sen, A. (1993), "Internal Consistency of Choice", *Econometrica*, 61, p. 495-521.

Schick, F. (1991), *Understanding Action*, Cambridge, Cambridge University Press.

Tversky, A. (1975), "A Critique of Expected Utility Theory: Descriptive and Normative Considerations", *Erkenntnis*, 9, p. 163–173

Verbeek, B. (2001), "Consequentialism, Rationality and the Relevant Description of Outcomes", *Economics and Philosophy*, 17, p. 181-205.

Weirich, P. (1986), "Expected Utility and Risk", *British Journal for the Philosophy of Science*, 37, p. 419-442.

Weirich, P. (2019), "Risk as a Consequence", *Topoi*, Special issue on "New Trends in Rational Choice Theory".