Hadrien Mamou

Rational preferences and reindividuation of relevant alternatives in decision theory: towards a theory of representation

Article (Published version)
(Refereed)

This version available at: http://eprints.lse.ac.uk/85905/

Available in LSE Research Online: March 2018

http://eprints.lse.ac.uk

CrossMark

Rational Preferences and Reindividuation of Relevant Alternatives in Decision Theory: Towards a Theory of Representation

Hadrien Mamou[1]

Abstract

In this essay, I will examine Broome's argument in *Weighing Goods* (1991; sections 5.4 and 5.5) that aims to show that moderate Humeanism, according to which any coherent sets of preferences should be rationally acceptable, is not a sustainable view of decision theory. I will focus more specifically on the argument Broome uses to support his claim, and show that although it may get some traction, it does not undermine moderate Humeanism as we know it. After reconstructing Broome's argument, I argue that standard decision theory supposes that the representation of a decision problem fully captures what is relevant to the agent, and therefore has nothing to say about the legitimacy of the representation in question. I argue that although Broome's criticism does not jeopardize moderate Humeanism, it does pave the way for a new area of research in decision theory.

Keywords Decision theory · Rationality · Preferences · Framing · Individuation · Broome · Savage

1 Introduction and Motivation

1.1 Introduction and Reconstruction of Broome's Argument

Standard decision theory provides an axiomatization of expected utility theory. Simply put, it shows that if an agent's preferences respect a series of axioms, then she can be said to be acting as if maximizing an expected utility function. These axioms are then usually considered as normative axioms of rationality, some of them being fairly uncontroversial: transitivity of preferences is for instance widely accepted in the literature. Given this axiomatic framework, different normative approaches can be taken with respect to rationality. One of them is commonly known as moderate Humeanism, introduced by Hume, (1740, p. 267), and

"Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger.'Tis not contrary to reason for me to chuse my total ruin,

to prevent the least uneasiness of an Indian or person wholly unknown to me.'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and have a more ardent affection for the former than the latter. In short, a passion must be accompani'd with some false judgement, in order to its being unreasonable; and even then'tis not the passion, properly speaking, which is unreasonable, but the judgement."

Moderate Humeanism thus claims that as long as some consistency constraints over judgments are respected, any desire-induced preferences should be acceptable: self-destructive preferences and cold-blooded pragmatism are just equally rational as long as they yield consistent preferences. Under this view, the axioms laid out by standard decision theory are considered prima facie as tentative reasonable requirements of rationality in that they capture intuitions of consistency as minimally as possible. Humeanism thus claims that all preferences, taken individually are rational; they only may not be so if they are not consistent with one another (or if the underlying desire is combined with a false belief). On the other hand, anti-Humeanism claims that there exist irrational preferences, even when considered in isolation. Although there is no uncontroversial indication of which classes of preferences are impermissible, as different theories push for

✉ Hadrien Mamou
 h.mamou@lse.ac.uk

[1] Department of Philosophy, Logic and Scientific Method, London School of Economics, Lakatos Building, Houghton Street, London WC2A 2AE, UK

🍀 Springer

different ones, anti-Humeans believe that some additional constraints should be imposed on preferences beyond those of standard axioms.

In *Weighing Goods* (1991; section 5.4 and 5.5) Broome provides a new argument against moderate Humeanism. He argues that, if no additional constraint of rationality is added, an uncontroversial axiom such as transitivity becomes a vacuous requirement: it does not constrain preferences *in any way*; therefore, Broome argues, moderate Humeanism collapses into a fully permissive theory of rationality, that allows any set of preferences: Extreme Humeanism. Extreme Humeanism being widely unattractive, Broome has shown by appeal to a reductio strategy that anti-Humeanism is the only plausible position.

I this paper, I will focus on the specific argument Broome uses to show that moderate Humeanism collapses into its extreme counterpart. To understand this argument, let us first define more precisely the relationship between the transitivity constraint and the reindividuation strategy. Transitivity imposes pairs of preferences to be consistent with one another: formally if O is the set of outcomes, then the consistency constraint demands that for all a, b in O ($a < b$ $b < c$) → ($a < c$). To give an example, if John prefers tiramisu to panna cotta, and panna cotta to cannoli, then John needs to prefer tiramisu to cannoli in order to have transitive preferences. Although this constraint is uncontroversial people may take issue with the way it applies to the objects of preferences. One famous example was given by Sen (1993): suppose you are invited to your friend's house, and the latter lets you choose between two desserts, having the one you left for herself. Suppose that when offered an orange and a big apple you pick the big apple; when offered an orange and a small apple you pick the orange; and when offered a small and a big apple, out of politeness you pick the small one. Your pattern of preferences is then said to be cyclical (big apple > orange > small apple > big apple) and therefore intransitive. Sen's objection consist in saying that this is not the way we should look at the outcomes, and that we should take into account politeness in their their description. Once correctly described your preferences are no longer intransitive, and your consistency is safe. This strategy of reindividuation of alternatives has been frequently used to explain and legitimately rationalize preferences that were prima facie deemed irrational. However, this is precisely this strategy that is under Broome's scrutiny.

Broome's argument can roughly be put as follows. Consider Maurice, who prefers staying home when given the choice between staying home and going to Rome; who prefers going to Rome when given the choice between Rome and mountaineering, and who prefers mountaineering when given the choice between staying home and mountaineering. Maurice has allegedly intransitive and

therefore irrational preferences. However, one could plausibly argue that after all, Maurice's preferences are not actually irrational since Maurice does not represent this situation through three but four alternatives: he distinguishes "mountaineering when given the choice between home and mountaineering" and "mountaineering when given the choice between Rome and mountaineering". Maurice justifies this distinction by arguing that he would not want to look like a coward by refusing mountaineering and staying at home, while he would enjoy looking cultured by choosing Rome over mountaineering. This redescription of the alternatives is formally called reindividuation of the outcomes: the set of outcomes is redescribed in a way that better captures the agent's attitudes, and may make a difference for the decision making process.

More formally, Maurice's alleged preferences in this case are intransitive: $H > R > M > H$, if {R, M, H} is the defined set of outcomes offered to Maurice, where R stands for "going to Rome"; M for "mountaineering" and H as "staying home". One may want to redescribe the set of alternatives as {R, M, H_r, H_m}; where H_r stands for "staying home when offered to go to Rome" and H_m for "staying home when offered to go mountaineering". This reindividuation "rationalizes Maurice's initial preferences" in the sense that it provides a redescription that makes his new preferences consistent: $H_r >' R >' M >' H_m$. We will call the first (respectively second) set of outcomes O (O'), and the first preference relation $<$ ($<'$).

While Broome prima facie suspends his judgement as to whether Maurice has rational preferences, he contends that unconditionally allowing for reindividuation of alternatives gives rise to the "emptiness problem": transitivity becomes an empty constraint, as any preferences over a set of outcomes can always be rationalized by a finer reindividuation of these outcomes. "*Where will it all end?*", he asks (Ibid, p. 101). Consequently, he argues that the only way to avoid this pitfall is to deny that any individuation of outcome is in general acceptable, and suggests two possible ways of doing so: the principle of individuation and the requirement of indifference. Being two distinct possible remedies, it does not matter if they are compatible or not. The principle of individuation claims that two outcomes are different if and only if they differ in a way that makes it rational to have a preference between them (which implies in Maurice's case that H_r and H_m are identical outcomes). The requirement of indifference states that two individuated outcomes must rationally be equally desirable if there is no way in which it makes it rational for them not to be equally desired (which implies that H_r and H_m are different outcomes but that Maurice should be indifferent between them). Although these principles are only two possible options to regulate preferences and individuation, Broome argues that moderate Humeanism could not accommodate any of these principles,

as it would contradict the fact that any set of preferences is rational as long as they are consistent:

> A Humean, on the other hand, cannot be happy with the conclusion that there must be rational principles of indifference. Such a principle denies that certain specific preferences are rational, which is something a Humean cannot allow. She cannot allow that rationality should ever deny a person the right to prefer anything to anything else, provided this preference is consistent with her other preferences. (Broome 1993, p. 5)

Broome thus infers that moderate Humeanism must accept that consistency constraints such as transitivity cannot apply to preferences. Consequently, moderate Humeanism is irreversibly exposed to the emptiness problem and collapses into Extreme Humeanism:

> A Humean must therefore pay the penalty. She will have to accept that rationality does not constrain practical preferences. [...] How severe is this penalty? Does it completely undermine the position of a moderate Humean? An extreme Humean believes that rationality allows a person to have any pattern of preferences whatsoever. Even a moderate Humean, we now see, has to believe that rationality allows a person to have any pattern of practical preferences whatsoever. [...] I conclude that the moderate Humean cannot sustain her position. She must either become extreme or cease to be a Humean. (Ibid, pp. 6–13)

To summarize this section, Broome's argument can be reconstructed as follows:

(i) Accepted claim: rational preferences are necessarily transitive.
(ii) Counterexample to the claim: Maurice has rational and intransitive preferences < over O.
(iii) But Maurice does not violate transitivity: we can reindividuate outcomes so that Maurice has non-intransitive preferences <' over the new outcomes in O'.
(iv) However, the emptiness problem arises: if any individuation is allowed, transitivity becomes an empty constraint, which is unacceptable.
(v) There are two ways of rationally constraining outcome individuation: either individuation is not always licensed; or indifference between outcomes is sometimes required.
(vi) But moderate Humeanism rejects both of these constraining strategies.
(vii) Therefore, moderate Humeanism collapses into extreme Humeanism.

I have so far exposed Broome's argument against Humeanism because of the "emptiness problem". In the next section, I will introduce the notion of relevant representations, make a claim about it, and argue that premise (vi) of Broome's reasoning is false.

1.2 Reindividuation and Representation: Literature Discussion

Besides Broome's address of the normative status of reindividuation strategies, few scholars have specifically discussed the issue. To the best of my knowledge, only two salient attempts have been made in this respect besides Broome's: Dreier (1996) and Pettit (1991). Pettit comments the problem of individuation upon his analysis of decision theory's ambition to be the "central sound core of psychology through idealization". He claims that "decision-theoretic probabilities and utilities which an agent would have under the idealized circumstances described in the relevance conditions are his beliefs and desires." (Ibid). If successful, "decision theory explicates certain platitudes about belief, desire, preference and choice". By contrast to Lewis (1983a), Pettit argues that this ambition is incompletely fulfilled, due to three main deficiencies. First, the author claims that decision theory is an incomplete account of the matters relevant to decision making, as it misses some of the elements that play a role in preference formation. In particular, he claims that by describing people's choices with attitudes to prospects (e.g. the prospect of "eating an small apple" versus that of "eating an orange" in Sen's example), it doesn't explain how properties realized by these prospects play a motivational role for agents' choices. As an example, suppose that John desires to go to the movie tonight and that this option would simultaneously disappoint his mother as she wants to spend time with him. The same prospect realizes two distinct properties that do not hold the same motivational content. This prospect is simultaneously desired and not desired by John. Yet claiming that the extensionality of preferences fails here would be misguided, because what truly drives our preferences are properties and not prospects according to Pettit. Based on this thesis, Pettit argues that decision theory is not an autonomous theory of rational idealized psychology, in that it needs some account of what identical outcomes are. Indeed, decision theory takes as inputs attitudes to outcomes (or prospects), without determining what counts as a such outcomes. Pettit goes on to say that the theory thus allows any individuation of outcomes. In other words, Pettit uses the problem of individuation to conclude that decision theory must depend on some principle of identification of outcomes, and believes that this principle should resort on properties rather than prospects. Moreover, this property-based principle must lean on the practice of folk

psychology, confirming the initial claim that decision theory is non-autonomous.

Although it does refer to the problem of individuation, Pettit's discussion is not centered on it as it only refers to it to bring forth his more general claims about the theory. More importantly, he does not discuss the logical compatibility of moderate humeanism with what Broome calls the emptiness problem, which constitutes one of the main concerns of this paper. Finally, I think that his conclusion that decision theory must rely on some folk psychology practice in order to solve the issue of individuation is not explicitly supported and therefore unconvincing. In this paper, I will in a way agree with Pettit on the incompleteness of the current theory, but offer arguments in favor of a solution to the problem individuation that is both related and independent from decision theory. My suggested theory is related to decision theory, as it is still idealized, normative and does not rely on psychology; yet I believe it should be normatively independent from decision theory in the results it produces, as I will explain later.

By contrast with Pettit, Dreier addresses the problem of individuation more frontally. Hoping to save moderate Humeanism from the emptiness problem, he provides an alternative solution to it. He first explains why Broome's position is intuitively unappealing and argues that by imposing intrinsic constraints on preferences it is overkill: it makes rationality too constraining, although Dreier believes that Maurice can rationally hold his preferences between R, H and M as long as he has reasons for it:

> We may ask Maurice, do you prefer hiking to staying home because you care about the relation the choice of hiking bears to the particular alternative of staying home? If he doesn't - if, for example, he simply hadn't noticed that his pairwise preferences among hiking, Rome, and home were intransitive, or if he did notice it but did not care - then after all his preferences are irrational. (ibid)

In other words, there is for Dreier a criterion that allows some Maurice-like preferences while avoiding the emptiness problem. And this criterion should assess whether Maurice's preferences are supported by another class of preferences. Indeed, if Maurice could imagine the choice between "Home when offered to go to Rome" and "Mountaineering", and that he did prefer the former over the latter (so that $M < H_r$), it would be sufficient to justify Maurice's rationality. This preference is akin to a thought experiment, as it cannot happen in real-life, but for Dreier it does provide a necessary and sufficient justification to Maurice's rationality. This is Maurice's specific contextual desires to the relation between home and mountaineering that makes the choice apparently cyclical but in fact rational: staying home holds specific properties when paired to mountaineering. Although I agree

with Dreier's intuition that Broome's solution is overconstraining, I think this new consistency condition is correlated with a rational justification, but neither necessary nor sufficient. Indeed, successfully showing that $M < H_r$ is not necessary nor sufficient to justify the rationality the initial cycle of choices $\{H < M < R < H\}$. To see this consider a case where Mauricette exhibits similar choices between H, R, and M, only this time, the feature that makes her choose this way is the special relationship between Rome and staying home, rather than that between mountaineering and staying home. Indeed, she chooses mountaineering over staying home as she loves skiing, Rome over mountaineering as she loves the city, but staying home over going to Rome because going to Italy when she knows she could have stayed home makes her home sick: she is very sensitive to the historical rivalry between Italy and her home country. Again, Mauricette allegedly has the same cyclical preferences as Maurice: $R < H < M < R$. Yet Dreier's justification for Maurice's preferences doesn't hold for Mauricette: nothing in this new story suggests that $M < H_r$: until proven otherwise, it seems that Mauricette prefers skiing to staying home here. Therefore, there are cases where it is not true that $M < H_r$, yet Mauricette's preferences still seem intuitively justified. Consequently, Dreier's solution does not identify in a necessary way the allegedly justifying consistency condition that saves the chooser's rationality.

A second remark about Dreier's claims: it is not sufficient that $M < H_r$ for it to justify the initial cycle either. Consider Alanis that goes by the exact same description as Mauricette, only this time she also happen to prefer H_r to M, if asked about it: she does love skiing, but not as much as she enjoys the relief of not being homesick in Rome. Alanis thus passes Dreier's test as $M < H_r$. Is it a sufficient justification to Alanis rationality? Not intuitively, as by Dreier's intuition we should add that $H_r < M$ in order to make this justification proper: what truly motivates and rationalizes Alanis cycle by Dreier's view should have been the peculiarity of the contextual desires between Rome and staying home. About both of these points, Dreier could amend his criterion and argue that all of these impractical relational preferences should be tested, and that we would eventually find out that $(H_r < H)$ but this condition becomes much more constraining: an agent should now check *all* of the potential relational (or alternative-dependent) preferences she has in order to claim her rationality. Moreover, it fails to explain cleanly what the relevant objects of choice are, and why such preferences occur. Finally, one can interpret Dreier's strategy as saving the moderate Humean view by expecting agents to make even more pairwise comparisons than before. If we treat Maurice's case (at least partially) as a philosophical disagreement between Maurice and the experimenter, this strategy boils down to shifting the burden of proof from the experimenter to Maurice: ex ante, on which ground can the

experimenter refuse Maurice's individuation? On the ground that this individuation is comes from abstract impractical preferences that justify it. In my view, accepting this shift of burden is implausible, especially in richer choice situations. Should Maurice always evaluate impractical pairs of alternatives before settling on an individuation? All in all, my point is that considering preferences over impractical thought experiments doesn't rigorously justify the agent's rationality, although it may sometimes be evidence of it. In my view, only a theory that indicates which objects are relevant to the agent choice and how she represents these objects would be satisfactory. Such a theory would have the merit of not being ad hoc by justifying the agent's rationality ex post, and would rather explain how such peculiar yet rational preferences arise.

2 Relevant Representations

2.1 The Relevance Claim About Decision-Theoretic Representations

The reader should first note that throughout the essay I will use interchangeably the words representation and description by convenience. Both terms are used in the (voluntarily general and intuitive) sense of some formalized attempts to portray the features of an actual situation of choice. The only important distinction here is between description and redescription, where the latter is only strictly defined as reindividuation of outcomes (cf above).

To theorize decision making, decision theory introduces primitive objects such as consequences, preferences and states of the world, and builds decision models on them. In Savage's framework for instance, actions are defined as mappings from states to consequences; where the states "must leave no relevant aspect undescribed" (Savage 1954, p. 9) so that all the features of the consequences that the decision maker cares about are represented. As an example, if I must choose between going to an Italian restaurant and staying home, I must include in my description of the decision problem that the Italian restaurant may or may not serve tiramisu, may or may not be closed, etc. insofar as I care about these features. This requirement exists both for technical and intuitive[1] reasons: in order to make accurate predictions and prescriptions, a theory of decision making must capture all that is likely to affect the satisfaction of the agent's ends.

I believe that this requirement has an important implication on the way we deal with decision problems in decision theory. In fact, *when given the description of a decision problem, we are also given a key assumption: the description is maximally relevant for the choice problem, as nothing more is needed (and available to the decision theorist/ adviser) in order to make a decision.* I will refer to this claim as the relevance claim. The relevance claim thus consists in assuming that every consequence, state, and option the agent cares about to make up her mind is is already included in the description of the decision problem. Intuitively, relevance thus has a similar object as a theoretical hypothesis,[2] that supposes the descriptive validity of a representation. To make this claim intuitive, consider a standard textbook example: you are to choose between different monetary bets contingent on the color of a ball drawn from an urn. The example here identifies the outcome space to lotteries of monetary payoffs. As it is, this description (that can be represented in a state-consequence matrix) presupposes that nothing relevant is left aside: in particular, all you care about is the financial outcome of each bet, and not, say, the esthetic pleasure you get from the color of the ball. Since if you did care about the color of the ball for its own sake, this should be included in the description of the consequences. Consequently, every representation of a decision problem has an attached implicit claim: that it is maximally specific as to what the agent cares about that may be related to the choice problem.[3] At least, this claim best explains the practice of decision theorist in the way the lay out their theory.

I will give two additional arguments that support this claim. First, if the description of the decision problem was not judged relevant, there would be no point in acting based on that description; I believe that it would in fact undermine the whole normative enterprise of decision theory. How can you possibly prescribe anything if you cannot trust some description of the problem that you are trying to solve? This is why, in decision theory textbooks, once a description of

---

[1] To see more on Savage's technical reasons see Bradley (2016, p. 9, Savage's Theory.

[2] As defined by Giere (1983), as an assumption about the relationship between a model and the real world. This comparison to Giere should only be understood as an intuitive analogy, as I am not necessarily committed to Giere's account of modelling here.

[3] Before defending this claim any further, I want to explain what I mean by maximally specific. Although we rarely think of decision representations this way, we hold the intuition that if the agent knew some relevant objects to be excluded of the representation, we should surely have them included too before making any normative considerations about the case. This suggest that there is a "more relevant" relation of order between representations, although it may be incomplete. Is there such a thing as a maximally relevant set? We know there is, as the set of all possible objects that may be included in the representation is (weakly) more relevant than any of its subsets. This doesn't mean that this maximal representation is the one we are after, as this superset may have many subsets that are themselves as relevant as this superset although they contain less objects. This subsets simply excluded irrelevant objects from the representation. All of these I will call maximally specific representations. For a more intuitive grasp on this, see the notion of excessive representation in the next subsection.

the decision problem is given, the theory usually gives a clear-cut assessment of the rational attitudes and decisions so that every relevant aspects of the problem are given. This assumption of relevance is what gives normative substance to decision theory as we know it. If a theory aims at evaluating the consistency of a set of choices, it has to assume an underlying set of attitudes to the world, AND suppose that these attitudes capture all that is relevant to the decision from the decision maker's perspective.

Second, if the representation is not supposed to be relevant, then the inconsistencies could very well come from the incorrectness of the representation itself: an agent (or an observer) could systematically blame the representation and save the agent's reasoning.

2.2 Implication of the Relevance Claim for Broome's Argument and Potential Objections

If we now grant the correctness of the relevance claim and of its underlying idea of maximal relevance, then Maurice's example can be interpreted in a quite different way from Broome's. The difficulty around Maurice's case lies in the fact that he might be recklessly using reindividuation to rationalize his attitudes, which may make the transitivity principle vacuous. However, we can now object that decision theory (as we know it) does not *allow nor disallow* reindividuation and thus reckless rationalization of agent's attitudes, but often has *nothing* to say about it: normative interpretations of representation theorems just tell you how to be consistent based on one given representation that is deemed relevant. Here Maurice gives two descriptions that are in turn claimed to be maximally relevant: one where considerations of awkwardness are excluded is ({R, M, H}) and one where they are included within the alternatives ({R, M, H_r, H_m}). But Maurice cannot simultaneously argue that both of them are maximally specific. At most, he could say that exactly one is maximally relevant, but it is not up to decision theory to decide which. Moreover, the logically impossible fact that both descriptions may be maximally relevant does not jeopardize transitivity and decision theory as we know it, simply because decision theory does not have anything to say about the choice of the relevant description, the conditions under which a description is relevant, or whether the agent should hold simultaneously consistent representations. From the first representation, decision theory will say that if it does capture everything Maurice cares about, then Maurice has intransitive and thus irrational preferences. And the theory will allow Maurice's attitudes and choice in the second representation; it is however, a *different Maurice*, in a *different decision problem* that the theory is assessing.

Moderate Humeanism, as Broome defines it, is a principle that applies to decision theoretic objects: preferences and outcomes. Consequently, this is a claim made at the level of decision theory: given a representation of the decision problem, moderate humeanism demands that the only constraints be consistency ones. Since Broome's objection regards reindividuation strategies and therefore choices of representations, it does not affect moderate Humeanism, since as Decision theory, moderate Humeanism has nothing to say about these. The upshot for Broome's argument is that step (vi) (moderate Humeanism rejects both of the constraining strategies established by Broome) is false: moderate Humeanism does not allow *nor* disallow to constrain reindividuation strategies. It thus does not make the theory empty, nor collapses into Extreme Humeanism. In a way, Broome's argument can be seen as category mistake, as it expects standard decision theory and moderate humeanism to constrain strategies that go beyond their scope.

I will now consider a series of objections against my view and attempt to defuse them. First, it could well be argued that decision theory should have something to say about relevance and the way agents represent their decision problems. But this would be a new field of investigation for decision theory, that would normatively question models of models, ie how we should model our representations of the world. My views can be summarized in the diagram page 11. Two phases of the process of decision making can be distinguished. The first one which is representational, and the second one about the assessment and resolution of the decision problem given a particular representation. The second process has nothing to say about the first and vice-versa. One may argue that since standard decision theory normatively engage with people's attitudes prima facie, it does have something to say about what is relevant and what is not. I believe that this objection is ungrounded. We do want to separate what the agent or modeller of the decision problem says to be relevant about it, from what standard decision theory prescribes doing and not doing. Why? Precisely because we want to be able to explain, understand and justify what is a bad decision given a correctly represented decision problem. This would not be possible if both problems were addressed simultaneously: either the representation could systematically be blamed, or worse, a bad representation and a bad decision theory could be licensed. To illustrate this point, consider a case of framing effect such as the Asian disease case. A disease is expected to kill 600 people. Subjects are being asked to make choices between two pairs of policies:

- First offer:
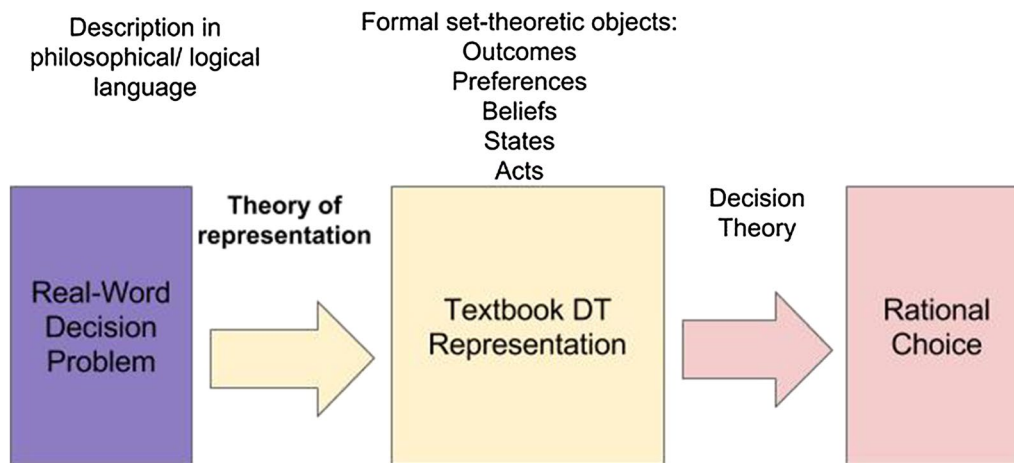   Program A: "200 people will be saved"

Diagram 1 Logical steps of the proposed framework of representation of decision problems and subsequent decision making

Program B: "there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved"
- Second offer:
   Program C: "400 people will die"
   Program D: "there is a 1/3 probability that nobody will die, and a 2/3 probability that 600 people will die"

In practice, people usually prefer A to B and D to C (Tversky and Kahneman 1981). Traditionally, the theory interpret these results as inconsistent, as programs A and C and B and B are respectively deemed identical, if we only consider the number of deaths and the probabilities. However, in order to claim that agents are inconsistent, we must suppose that the two descriptions represent the same (or equivalent) decision problems. In particular, can we say that agents are inconsistent if "not die" and "be saved" are not equivalent descriptions? Is it the case that the subject only cares about deaths and lives as consequences?

I thus suggest that the process of decision theory and advising be separated in two independent fields (see Diagram 1): Starting from a description of the agents' attitudes about the real world decision problem in non decision theoretic language (in a logical language or one of philosophy of action for instance), a theory of representation would then prescribe how to describe the problem in decision (set-) theoretic terms and which of the possible representations are admissible ones. Only then can standard decision theory be applied to the selected representation. With these two frameworks in mind, it seems that the scope of Broome's argument about how to constrain reindividuation and how to solve the emptiness problem covers only the first step of the decision making process, namely determining the relevant representation(s) of a decision problem, and more importantly the licensed methods to reach such a description.

A second potential objection can be aimed at this division of labor between the relevance of the representation and and the rationality of the decision making. One could argue that in many cases, decision theory does tell you about what is relevant or not for the decision. Consider for instance the textbook urn case given earlier: suppose the agent also cares about the room temperature when she draws the ball, as she prefers to be warm rather than to be cold. However, the room temperature is always the same, so that the outcomes only differ in the color of the drawn ball. Then indeed, decision theory tells you that the temperature of the room should not play a role in the agent's decision, that it is not relevant in some sense. Although normative principles of decision theory yield some constraint on the representation, the nature of the constraint is structurally different from the one Broome has in mind. In order to make this point clear, I will first refine the notion of representational relevance.

A description of a decision problem can be bad for several reasons. It may be bad because it insufficiently or illogically describes reality (which may be the world, or the agent's relation to the world). The description can also be bad because it is excessive (as in the room temperature case). Now, when we say that the representation should capture everything that is relevant (and possibly more), we mean that descriptions should be sufficiently exhaustive. But it can be the case that they be logically inconsistent, or excessively precise. When a representation of a decision problem is excessive or logically inconsistent, then indeed, decision theory may show that it is the case. But precisely because its principles are consistency principles, it will never tell you that the representation is missing something that is relevant to the agent (in the sense that the agent cares about this thing). We can thus refine the previous framework, by

saying that representation theory should establish which set theoretic descriptions are sufficiently accurate, although they may be excessively or inconsistently so; and decision theory may sometimes tell you which features are superfluous for decisions (but I am not sure that it always does.). And again what would happen if we mixed the two fields? We would not be able to check whether the relevance principle stands. Imagine a DM who knows about what matters to him, but is mistaken in his subsequent deliberation. A theory of decision-making that would simultaneously assess objective and subjective relevance, would not be able to identify the source of mistake. Second, the opposite problem of identification can occur if the agent is not rationally blameworthy but initially misrepresented the decision problem.

2.3 Is Moderate Humeanism Any Safer Under This View Than Under Broome's?

A more general objection has been raised about the status of moderate humeanism and the emptiness problem before representation this time. If I understand the argument correctly, it objects that although Humeanism may be safe at the decision theoretic level, it may not be at the pre-decision theoretic one: what guarantees that all preferences should be acceptable as long as they are consistent, and so before representation? As I understand the Humean versus anti-Humean debate, the main bone of contention concerns the normative legitimacy of practical reason. Consequently I will leave aside positions about moral or legal normativity. In particular, the moral status of ends, means, and desires are off the scope here. The relevant part of the objection is whether all preferences should be rationally acceptable, even before representation, as long as they cohere. This paper does not address the worry. The reader might thus wonder how my response to Broome solves the problem for moderate Humeanism here. In this paper, my hopes were twofold. First, to argue that Broome's ambition fails, as he *does not successfully* show that the formal decision theoretic apparatus suffices to prove moderate humeanism wrong. Second, I advocated the establishment of an independent upstream theory, that would ask the normative and more semantic question of representation before decision. True, this does not prevent a general anti Humean argument to go through at this upstream level. *But I would not have hoped to achieve so much in one paper*. Yet it is very likely that some primitive objects of an adequate theory of representation will not all

be identical to Savage's. The notions of care about objects of deliberation, and of relevant features for choice cannot be captured by the theory so far, unlike (arguably) attitudes of desire and belief. Consequently, care and relevance in decision making should be reduced to (or introduced as) at least some new primitive attitude. And I suspect that such considerations will only be settled once the question of representation will be.

2.4 Broome's Argument Within Jeffrey's Decision Theory

In what follows, I want to consider here a more technical objection. So far my claims as well as Broome's were formulated within the standard Savage framework of decision theory. The present objection involves Jeffrey's theory of decision making, which I will describe now. As explained above, Savage's theory shapes decision making along three intuitive dimensions: states, beliefs and desires. This representation strategy is double edged, as although it offers nice categories of decision making, it also constrains the way a real life decision can be represented. In particular, one should be able to strictly distinguish the object of beliefs (states), of desires (outcomes), and of means to bring about these outcomes (options). The theory implies further that the desirability of an outcome ought to be independent from the state of the world where it occurs (state neutrality), as well as from the subjective probability state in which they occur. Finally, the options themselves should not impact the probability of the various states of the world involved in the decision problem. By contrast with Savage's, Jeffrey's theory is puzzlingly flexible, as it presupposes neither these conceptual categories of decision making nor the constraints that go with them. Objects of desires, beliefs, and choices are all propositions, and thus possibly interdependent for Jeffrey. As an example, the fact that it rains tomorrow can simultaneously be an object of desire and belief for Jeffrey, but not for Savage. This structural flexibility allows decision theorists to model real-world decision without any reformatting of the options, states and consequences, but rather as the decision maker perceives them. This leads us to the second major virtue of Jeffrey's theory, namely that it can afford describing the various acts, options, and states in a dependent or independent way, such that the theory will always prescribe the same choices. Unlike Savage's theory that only allows for independently described objects, Jeffrey's theory is partition invariant: any description of the decision problem through any partition of the state space will yield the same result. This property is best understood formally through the partition of a proposition p: $\{p_i\}$ is a set of mutually incompatible but jointly exhaustive sub-propositions by which proposition p can be realised. $p = U_i\{p_i\}$ and if the agent's preferences respect Jeffrey's conditions

of rationality, they can be represented through a desirability function Des(p), very similar to expected utility but applied to any p : $Des(p) = \sum_i Des(p_i).P(p_i|p)$. Given this result, the partition invariance property is guaranteed by the fact that an agent respecting Jeffrey's axioms of preferences will assign the same desirability to p, regardless of the partition $\{p_i\}$ in the above formula. Partition invariance is normatively attractive (see for instance Joyce 1999), as "it ensures that various representations of the same decision problem yield solutions that agree." (Weirich 2016).

How does it relate to the problem of individuation and my present claims? Since I advocate a theory of representation on the ground that decision theory doesn't say anything about it, and that rational decisions are sensitive to these representations, one could object that this concern only applies to Savage's theory. Moreover, since Jeffrey's theory is partition invariant, shouldn't it yield the same choice in the two Maurice's cases? To put it differently, a Jeffrey agent would allegedly be consistent across representations, so why would the relevance claim that I defend still be substantial? Note first that if it were the case, my answer to Broome wouldn't go through in this alternative framework, all the more so as Broome's worry would itself be defused at once: according to this objection, either Maurice is respecting Jeffrey's axioms and chooses consistently across the board, in which case Humeanism is safe, or he doesn't and is rationally blameworthy, and Humeanism is safe too. To put it differently, this objection suggests that the emptiness problem doesn't exist in Jeffrey's framework, as reindividuation shouldn't change anything about Maurice's choice.

To assess this challenge, let's see how Maurice's situation can be translated in Jeffrey's framework. In the initial situation, Maurice also exhibits cyclical preferences with respect to the three propositions "going to Rome", "going mountaineering" and "staying home" (respectively R, M, H). In the second, redescribed situation, Maurice has the following preferences: H&notC > R > M > H&C; where C is the proposition "looking like a coward". So far, the DT descriptions looks very similar to Savage's. But there is more to it: since Jeffrey's objects of preferences are propositions, they entertain logical relations with one another. One way to see the structure of propositional relations in Jeffrey is through the desirability equation mentioned above: if $\{p_i\}$ is a partition of proposition p, the desirability of p is then given by the following formula, that gives a weighted average of p's desirability: $Des(p) = \sum_i Des(p_i).P(p_i|p)$. Applied to Maurice, if we take {H&C, H&notC} to be such a partition of H, and p to be H:

$Des(H) = Des(H\& notC).P(notC/H) + Des(H\& C).P(C/H)$;

this is simply due to the fact that $P(H\& notC/H) = P(notC/H)$.

Therefore, the desirability of H exists, and is defined as some weighted average of the desirability of the finer propositions that capture the cowardice consideration. Consequently, Jeffrey's framework imposes that H and C be such that H&notC > H > H&C. The difference of desirability between H&notC and H will be more or less significant, depending on the probabilities of C and notC, and this will also impact the preference between R and H, and R and M.

Now that the translation in Jeffrey's term is established, what does it say about Maurice's rationality? Well, nothing. As the story told in Savage's case would go exactly the same way: Maurice is implicitly asked about H, C and R, and then argues that H didn't actually capture accurately the choice he was facing, so that what he actually evaluated was H&C and H&notC. As before Maurice cares about more than the mere destination. If the choice is explicitly made by another agent in propositionalized term, one could say that Maurice ought to exactly understand the offer at its face value, ie that by "staying home" nothing more is meant than staying home, whether it implies looking like a coward or not. If he did understand it this way he would be expected to choose in a transitive way, would argue the objector. However, I don't think that this requirement of understanding the offered choice at its face value is a requirement of rationality: it may be a semantic requirement, a requirement of understanding and cognitive ability, but it is not what anyone would expect from a rational agent. I think one could imagine a rational agent that needs to be specified exactly the scope of the offered alternatives anytime she is about to make a choice. Granted it would be a weird individual, but intuitively not an irrational one. Therefore, moderate Humeanism is not concerned with these kinds of requirements, as they are not rationality requirements, and again I don't think that Broome's argument follows through. This doesn't mean that nothing needs to be done, as a theory of representation would be just as useful in Jeffrey's case. As the objection only made sense if the offered was made in propositional term by another agent; but if the story only is about Maurice considering his alternatives and choosing, then the question of how to represent properly the alternatives he is facing comes back as powerfully as before. And then, the issue is not linguistic, but representational: based on which rules should Maurice include specific features of the alternatives in his decision problem in order to apply decision theory to it?

3 Conclusion

In what precedes, I reconstructed Broome's argument against moderate Humeanism and show that decision theory as we know it has nothing to say about about the relevance of the representation of a decision problem. Consequently, Broome's argument about the emptiness problem and licensed reindividuation only regards the methods of choice of representations. Since moderate Humeanism as we defined it only regards the way decision theory (ie the second step of the normative evaluation) is done, it is not undermined by Broome's argument. However, the core of his argument may still very much apply to a new area of research, that may investigate what a relevant representation should look like.

Compliance with Ethical Standards

Conflict of interest This statement is to certify that all Authors have seen and approved the manuscript being submitted. We warrant that the article is the Author's original work. We warrant that the article has not received prior publication and is not under consideration for publication elsewhere. On behalf of all Co-Authors, the corresponding Author shall bear full responsibility for the submission. This research has not been submitted for publication nor has it been published in whole or in part elsewhere. We attest to the fact that all Authors listed on the title page have contributed significantly to the work, have read the manuscript, attest to the validity and legitimacy of the data and its interpretation, and agree to its submission to the Journal Topoi. All authors agree that author list is correct in its content and order and that no modification to the author list can be made without the formal approval of the Editor-in-Chief, and all authors accept that the Editor-in-Chief's decisions over acceptance or rejection or in the event of any breach of the Principles of Ethical Publishing in the Journal of Topoi being discovered of retraction are final. No additional authors will be added post submission, unless editors receive agreement from all authors and detailed information is supplied as to why the author list should be amended.

References

Bradley R (2016) Decision theory with a human face. Available via http://personal.lse.ac.uk/bradleyr/book.html. Last modified 25 Nov 2015

Broome J (1991) Weighing Goods. Wiley-Blackwell, Cambridge

Broome J (1993) Value, welfare and morality. Cambridge University Press, Cambridge

Dreier J (1996) Rational preference: decision theory as a theory of practical rationality. Theor Decis 40(3):249–276

Giere RN (1983) Testing theoretical hypotheses. Minn Stud Phil Sci 10:269–298

Hume D (1740) A treatise of human nature. Oxford University Press, Oxford

Joyce J (1999) The foundations of causal decision theory. Cambridge University Press, Cambridge

Lewis D (1983a) Philosophical papers, vol 1. Oxford University Press, Oxford

Pettit P (1991) Decision theory and folk psychology. In: Bacharach M, Hurley S (eds) Foundations of decision theory: issues and advances. Blackwell, Cambridge

Savage L (1954) The foundation of statistics. Wiley-Blackwell, Cambridge

Sen A (1993) Internal consistency of choice. Econometrica 61(3):495–521

Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211(4481):453–458

Weirich P (2016) Causal decision theory: Stanford encyclopedia of philosophy