ESJP
#5 | 2013

Getting the Description Right

Saving Transitivity from Arbitrary Translation

Darian Heim

1. Introduction

Stories with a good pinch of irony often reveal how and where an ideal is caught up by reality. We begin this paper about decision theory with such an example: Jane, a scholar in this discipline, goes to her favorite teahouse to have her 4 o'clock tea and a piece of cake as always. The waiter offers her the choice of the day: 'Today we have on our menu: apple tart, blueberry muffin, and cheese cake.' Jane decides for the first, apple tart. However, after a couple of minutes, the waiter comes back to tell her that he had forgotten to mention that there is lemon pie, too. '*In that case*,' Jane answers after a moment, 'I'll take the cheese cake[1].'

Although the punchlines of well-told anecdotes are supposed to be self-evident, I hope the reader will excuse my insistence to elaborate. We can infer from Jane choosing apple tart (*a*) among the initial set of alternative choices, that is, apple tart (*a*), blueberry muffin (*b*), or cheese cake (*c*), that she prefers *a* over the other choices. The introduction of a new option in the set, the lemon cake in this case, should not overturn Jane's initial choice a unless the new option is preferred to *a* itself. However, this is not what happened: we would surely not expect Jane to choose *c* given the preference she revealed first (*a* over either *b* or *c*). This is the ideal of what is considered rational behavior, namely, respecting transitivity in choice behavior – and the punchline of the above mentioned anecdote suggests that decision theorists, who deal with this ideal professionally, have the least confidence in it.

And yet the cake anecdote (henceforth CA) is relevant beyond the definite boundaries of a specific academic discipline. In everyday life, it is not uncommon to change one's opinion within different timespans and as a function of the information at one's disposal. For instance, Jane might – intuitively, perhaps – change her opinion on whether to take an umbrella with her on a cloudy day, what party or politician to vote for, or what partner to spend her life with (if this is a choice she is considering at all). In such contexts, it is not counter-intuitive that Jane attributes different utilities to the possible outcomes. That is, she she asks herself which situation, that would result from her action, she would prefer more that its alternatives given what she knows now about the weather, politics, or her partner. In doing so, she could eventually rank the different implications of her choices in a preference ordering. With regard to the second example, she might prefer to live in a quite egalitarian society rather than in a liberal welfare state which in turn she prefers to a libertarian meritocratic system. As a consequence, Jane would probably vote for parties on the left, rather than the right, end of the political spectrum. Moreover, we would expect Jane to stick to her voting behavior in similar conditions *given* her preference ordering. So, where does intransitivity enter the picture such that even non-adepts of decision theory might enjoy the punchline of our CA?

Imagine that Jane changes her political view and votes for a conservative party. The mere observation of Jane's changed voting behavior will appear inconsistent and thus intransitive to us. It would appear intransitive because her initial preference ordering concerning political systems is overthrown, and the initially least preferred option is suddenly chosen (and thus factually preferred). Obviously, one might say, context matters and her preferences are not unalterable. Jane's life circumstances might have changed. Imagine for instance that she has got tenure at a prestigious institution. She now feels for the first time the disincentivizing impact that progressive taxation has on labour and adapts her voting behavior

accordingly. However, the underlying question still stands: how are we to understand Jane's change of opinion, utility ranking, and preference from an outside perspective – while we are unfamiliar with her personal conditions? To illustrate this, suppose an old friend of Jane – who has not seen her in a while – gets to know what party she voted for. It is not too far-fetched that Jane's behavior (given her initial preferences) appears irrational from his perspective. The point here is that the decision theorist (or any other social scientist) is in a position similar to that of Jane's friend: certain givens and details about Jane are not available to him or simply not taken into account. The crux for both, old friends and scientists alike, is how to deal with preference changes during a period of interest for personal (the old friend) or scientific (the social scientist) purposes. More generally, how should we accommodate the fact that people (and their preferences) change over time? It is this difficulty that the initial CA illustrates and condenses into a preference change that happened within a couple of minutes. The underlying problem of both examples is the same.

Am I making too much fuss about an unavoidable fact of human existence? Jane, it seems, changed her opinions and preferences for the available options in view of new information or altered life circumstances. In fact, this is precisely what I want to argue for in this paper: Jane's changed behavior is not necessarily irrational and an effort should be made to interpret and consider seemingly intransitive changes in behavior in accordance with our theory of rational choice. Its rationality can be preserved by refining our – or the decision theorist's – assessment of Jane's specific choice and her underlying motivational considerations[2]. The strategy that I suggest to 'repair' the irrationality of Jane's choice for *c* is to describe the choices more accurately, that is, according to their specific context. Basically, we need to acknowledge that various factors could explain Jane's seemingly intransitive choice, for instance that her new job changed her political preferences. With regard to the initial CA, she might have had reliable reasons to believe that this would be the last time that *c* was available – after years of patronage of this teahouse, she knows that always when a new dessert option is offered for the first time (now, the lemon cake) the option that has featured on the menu the longest (*c* in this case) won't be served anymore the next day. Obviously, these aspects change the preference-relevant nature of option c such that it can be re-described as *c**, where the asterisk indicates the (new) knowledge of the cheese cake being offered for the last time. Although her preference ordering was *a* over *b* over *c*, we can now adopt an updated ordering in light of the new information without a violation of the principle of transitivity: *c** over *a* over *b*.

Thus, a refinement of the choice-description seems required in order to assess the rationality of an action properly. However, this strategy – as uncontroversial as it might seem – is not without problems. According to decision theorist and economist Paul Anand (1990), the re-description of a choice according to a specific context is arbitrary and leads to an insurmountable paradox. The issue is this: reasoning along the lines of re-description can warrant intransitive conclusions of *prima facie* transitive choice behavior too. For instance, if Jane abided with *a* instead of taking *c*, then the choice would be transitive according to our observation. However, if she was supposed to know about *c* becoming *c** at that point (and we, from the outside, knew about the underlying relevant information), then her ultimate choice for *a* would nevertheless be intransitive because it did not respect the eventual preference ordering *c** over *a* and *a* over *b*. Obviously, our conclusion depends on a clear definition (and thus description) of the specific choice options (including the type of information taken into account) and their preference ordering *before* we assess the rationality of her choice. However, we cannot define the set of variables that underlie and inform the choice description while also allowing for re-descriptions of choices according to an indefinite amount of additional factors. For Anand, the absence of a clear prior definition of these elements makes it impossible to give a conclusive assessment of the rationality of a choice – an initially irrational choice might become rational (and vice versa) in light of evidence we discover long after the choice was made. Re-descriptions are arbitrary: fundamentally, there are no clear boundaries or criteria to decide on what counts as a legitimate re-description.

In view of such arbitrariness, Anand (1990) rejects the possibility of re-descriptions to accommodate seemingly intransitive choices. In section 2, where his overall position is described in greater detail, we will see that his argument is based on a general critical attitude towards transitivity. In section 3, I shall indicate pathways for preserving transitivity while

contesting Anand's rejection of re-description based on two considerations. Firstly, his position depends on a division between descriptive and normative decision theory, which I deem unrealistic. Secondly, his position presupposes that observable choice and revealed preference are the only criteria on the basis of which options can be described. This, in turn, undermines the possibility of preserving transitivity by describing choice in terms of motivationally salient and potentially unobservable considerations of the agent. Therefore, I will argue that descriptions that include these terms are not necessarily arbitrary. In section 4, finally, the findings of this paper are briefly summarized.

2. Is there any normative appeal to transitivity?

2.1 The psychological justification

In his 1990 paper[3], Anand mainly focuses on the normative importance and legitimacy of the axiomatic method in the social sciences. Axioms allow us to model and imitate the relevant factors of real world behavior. They define the nature of the objects of study as well as the relationships among them. Anand emphasizes two definitional elements of the axiomatic method. The first element is what he calls 'choice primitives' – the objects or outcomes which are compared and assessed according to the axioms. In our example these are the types of dessert choice, represented by *a*, *b*, *c*, or *c**. The second element of the axiomatic method consists of the nature of the preference relation among these choice primitives – determining, for instance, whether they have to be ordered in a transitive manner or not. What matters for our purposes is that Anand is rather critical of the way axioms are used and justified in concrete scientific practice: he claims that both elements of the axiomatic method are unfounded, at best '*only partially* (or *un-*)*interpreted*' (Anand, 1990: 91, italics in the original)[4], and hence underdeveloped. According to him, in order for the model – which is ultimately defined by its axioms – to have normative appeal, the axioms need to be justified separately. The nature of the choice primitives, their characteristics and the stipulated relations among them need to be relevant and legitimate on their own. By what criterion can we determine whether the information we consider in our choices is the best candidate? How can we test whether transitivity is the optimal manner to judge the rationality of preferences?

Anand himself is critical of the normative appeal of transitivity. His attitude and implicit motivation[5] should be seen in the context of the advent of the experimental approach in the social sciences (predominantly in economics). Historically, Maurice Allais (1953) was one of the first social scientists to test axioms empirically. He provided empirical evidence that undermined some classical requirements of rational choice theory such as transitivity. What he observed was a 'certainty bias': if Jane – assuming that she was a representative participant in Allais' experiments – displayed a (slight) preference of *a* over *b* and if she was offered *b* for sure and *a* with a (high) probability she would choose *b*. However, in an analogous case, if *b* and *a* were both unsure outcomes and if the difference in probability between them was the same as in the first case, it would turn out that Jane would have chosen *a*[6]. Such a result violates transitivity because the preference relation between *a* and *b* changes although the preference-relevant information, that is, the difference in probability of the outcomes, stays the same[7]. The upshot of this critical tendency towards transitivity is illustrated best with a quote by Peter C. Fisher (Fisher, 1991: 29): 'The sanctity of transitivity as a bulwark of rationality and order will gradually erode, but this will take time.'[8]

This is the wider context of the empirical challenge brought forward against transitivity. In light of such findings, one might consider whether axioms like transitivity could be maintained on normative grounds instead – asking whether agents *should* act according to the axioms. Addressing this question in his 1990 paper, Anand discusses the '*Psychological View of Normative Appeal* (PVNA)' of the axiomatic method. In his view,

> [...] we should employ axioms because they describe propositions which we could easily accept. (Anand, 1991: 93, emphasis added)

Axioms like transitivity seem to reflect an intuition we have about rationality: it seems counterintuitive to consider a genuinely inconsistent (for our purposes this implies: intransitive) choice as rational. Behavior that

consistently violates any ordered and reconstructible pattern does not seem to warrant attributing rationality to it. Traceability and predictability go along with our common concept of rationality – we want to understand the underlying reasoning that motivated a choice before we call it rational[9]. However, whether a specific choice is transitive or not depends crucially on the specific way this choice is framed, expressed, and described. Ultimately, any difference in the description of choice behavior will have an impact on its normative content too, just like the distinction made between *c* and *c** in Jane's choice did. The more convincing the underlying 'story' about the construction and description of the primitives is, the more normative appeal the latter have.

Besides the arbitrariness of descriptions I alluded to in the introduction, there is a more fundamental problem, according to Anand. In the PVNA justification of axioms, the description of the primitives depends on its specific wording – its 'intensionality' (Anand, 1990: 93). However, it undermines the reason the axiomatic approach is used in the first place: we postulate and use simplifying axioms in order to *deductively* define rationality and thereby account for a multitude of concrete occurrences. Axioms like transitivity serve as the premises of the deductive argument. And from their truth follows the truth of the consequence, that is, the attribute of being rational for *all* situations where the axioms hold. Such an argument, however, is 'extensional' (Anand, 1990: 93) and independent of specific wording. To justify the use of axioms in terms of their PVNA, then, undermines this very deductive advantage. Every specific application of the axiom will have to be assessed separately in terms of accuracy of description. To use the words of Anand:

> While PNVA gives weight to cognitive factors […] it provides no grounds for the extensionality on which the use of axioms is based, and can only be regarded as ultimately self-defeating. (Anand, 1990: 93)

I wonder whether this is not too harsh and categorical a conclusion. Although the elaboration of my criticism with regard to his conclusion shall only begin at the very end of this section, we may already state that Anand's evaluative framework is clear-cut and strict: axioms have to apply independent of context or language. For this reason, it is necessary to define the domain and primitives to which the axioms apply before we assess specific situations. This, however, excludes re-descriptions *after the fact*, that is, once the behavior or choice in question has already taken place.

2.2 The translation theorem

In the further development of Anand's argument, he presents a formal proof of the arbitrariness of (re-)descriptions on the basis of what he calls the 'translation theorem':

> All intransitive behaviours can be redescribed in such a way the transitivity is not violated and all transitive behaviours can be redescribed in such a way that transitivity is violated. (Anand, 1990: 94)

The proof basically formalizes our initial example of arbitrary re-description. Whether Jane violates transitivity or not, whether she acts irrationally or not, depends on the description of the choice primitives one chooses. The theorem states that the (re-)description does not depend on the behavior itself: Jane's choice *c* might be 'translated' into any other primitive *c**, *c^*, etc. depending on the context, information, and hence description of the choice taken into account. But, Anand asks, is it legitimate to look for a 'better' description of Jane's choice after she has taken her decision already, that is, after the fact?

This is where the arbitrariness of re-description resides:

> Without prior agreement on the linguistic conventions which will be used to say what counts as a particular choice primitive, we can choose, *ex post facto*, some convention (richness of language permitting) in such a way that an observation (set) can be counted, either as a violation of transitivity or any other axiom [footnote suppressed] which we are testing, or not, depending on choice. (Anand, 1990: 96)

According to Anand, this arbitrariness is problematic. If we do not possess a definition of the set of choice primitives before we apply it to a concrete case, contradictions or an ad hoc theory will result. This is intuitively

plausible: in assessing a choice situation normatively (asking 'what is the rational thing to do for Jane?'), I must know beforehand what choices Jane might face and what information she has about them – adding or 'inventing' choices, or descriptions thereof, after the fact seems a patching ad hoc measure violating the requirement of completeness. Indeed, completeness is a precondition for transitivity[10]: without presuming that all possible choices are known, without a complete set of primitives, any discussion of transitivity is senseless as another choice may always 'pop up' and intervene with the established preference ordering.

Is there no way to avoid this problem? Anand mentions a 'consequentialist' option to deal with it. On this account, the choice primitives are defined with reference to what the agent perceives as her choice. Anand summarizes such a position in the following manner:

> [...] if the agent derives utility over something – if it is of "concern" to the individual as Savage put it – then, whatever it is, it should be modelled as part of the utility-yielding primitive. (Anand, 1990: 97)

Likewise, with regard to the example of Jane, we could say that whatever the preference-relevant context, consideration, or information, it has to be reflected in the way we construct the choice primitives. But, if any yielded utility to the agent had to be incorporated into the description of the choice primitives, then it would be impossible to assess whether the axioms are respected or not, whether the agent acted rationally according to those axioms or not. This is what Anand was interested in in the first place, testing axioms or assessing their normative appeal (cf. Anand, 1990: 97). If the agent was to have, say, a deliberate preference for intransitivity, it would of course be pointless to try to assess the transitivity of the actions. If Jane had chosen c only because she derived utility from deviating from predictions by rational choice theory, what sense would it make to ask – with regard to the axioms we postulate – whether this was a rational choice or not? For Anand, basing the description of the choice primitives on the agent's internal state is ultimately incompatible with normative assessment of choices, and with the very project of axiom testing.

2.3 Two objectives of decision theory

Accounts like the one above are thus in danger of producing absurd implications because they endorse 'pre-axiomatic proposals for primitive constructing' (Anand, 1990: 97). If the agent's utility determines the structure of the axioms, then the very reason to use axioms is subverted. Although theoretically possible, such a consequentialist solution to the arbitrariness-charge would be 'rather costly' in terms of

> [...] effects on the objectives which decision theorists set out to achieve (in particular the aims of providing a theory with behavioural content and one which is "hands-off" [...] (Anand, 1990: 97)

Now, what does Anand mean with 'behavioural content' and "hands-off"? This distinction is crucial, since Anand argues against the compatibility of the two objectives that are pursued in decision theory: the descriptive and the normative objective. The underlying idea is this. In decision theory, we are aiming at, on the one hand, a *descriptive* account of the ways people eventually make their decisions on a factual level. And such an account needs to simplify, explain or even predict the concrete behavior in question – it has to have 'behavioural content'. According to this objective, the decisions theorist aims to obtain an appropriate and realistic description of the choice – and the previous discussion of Anand's translation theorem applies to such an approach. On the other hand, however, we are looking for an independent and convenient tool or benchmark against which actual behavior can be *normatively* assessed without the decision theorist having to look at every specific case ('hands-off'). On this account, the re-description of choices does not need further argument, as the primitives need to be determined upfront. Evidently, Anand's criticism of re-descriptions after the fact presupposes this normative objective of decision theory.

So, what is the relation between the two objectives or approaches? Interestingly, with regard to the differences in primitive construction, Anand states on the one hand:

> Solving *simultaneously* for the twin objectives of being "hands-off" and providing content is akin to squaring the circle. (Anand, 1990: 100)

However, he seems to make a concession which is at odds with his own argument: 'Both projects seem to be reasonable ones to follow…' (Anand, 1990: 99). It is hard to see how the descriptive objective is supposed to be 'reasonable' in light of the challenge the translation theorem poses to it. Or, if the project can be somehow reasonable, will there be a criterion that ensures that (re-)descriptions are not arbitrary? Although Anand is not explicit about that, I want to argue for the affirmative answer to this question in the following section.

3. Getting the 'hands-on' description right

3.1 The interdependence of the two objectives

Let us start with a fundamental question straight away: why do we talk about the concrete normative appeal of axioms when they are artificially posited with the intention to serve a purely deductive purpose? The discussion of the axiomatic method is basically motivated by the observation of recurrent *prima facie* violations of those very axioms. The experimental results, presented first by Allais and subsequently widely confirmed, speak clearly against the axiom of transitivity in preference orderings (as displayed through choices). But intuitively, it is doubtful whether agents simply act irrationally on a systematic and widespread scale and thereby violate an axiom as important as transitivity. So, if we do not want to bite the bullet that humans consistently act irrationally, two different strategies are open to us: either we check whether rationality – as defined by the posited axioms – might apply to the undertaken actions in another manner, or, we ask whether those very axioms do indeed represent the best way for normatively assessing our behavior. I will defend the former strategy in the following pages, whereas Anand emphasizes the latter approach to axiom testing.

Experimental evidence is descriptive and hence inductive; it is based on concrete outcomes or choices. As such, it is brought forward against the normative conclusions of an axiom-based deductive approach. The point is that Anand's discussion of how axioms should be tested presupposes the interdependence of the two projects, the descriptive-inductive and the normative-deductive. It does so to the extent that it takes evidence from the empirical approach as a motivation and justification to scrutinize the validity of the axiomatic approach. Although Anand's two objectives are ultimately connected, he presents them as two independent goals. However, it is insufficient to declare a seemingly axiom-violating behavior as irrational merely on the basis of a *prima facie* observation: to call Jane's choice *c* straightforwardly irrational simply because she preferred *a* over *c* at would undermine the requirements of the descriptive approach. By barring any options that would preserve rationality, such a conclusion about Jane's behavior would be premature, as it would not consider the possibility that she has acted rationally after all (in view of information unknown to an outside observer, for instance). But the central question remains unanswered: of what nature is the relation between the two objectives in decision theory, if such a relation exists?

Anand argues that the two objectives are mutually exclusive approaches to decision-making processes (Anand, 1990: 92) – one pursues either the descriptive or the normative objective, but it is impossible to pursue both at the same time or only partially. As a result, to simultaneously solve both objectives literally amounts to 'squaring the circle' (Anand, 1990: 100) – it would be tautologically false and hence contradictory. Such a position, however, undermines the interdependence of the objectives and thus the very reason for testing axioms in the first place. Doing so begs the question of Anand's own project. But it is not my intention to enter the century-old discussion of induction and deduction here. The upshot is that Anand's dichotomy is too strict, too rigid, and deeply unrealistic with regard to any concrete decision-theoretic take on choice behavior. What I mean with this claim shall be shown by returning to Anand's translation theorem.

3.2 The Translation Theorem revisited

According to Anand, choice primitives should not be described *ex post*. Instead, we should follow the normative objective and determine our 'linguistic conventions' with respect to the choice primitives prior to the observation of the choice. What we can postulate about the decision-mak-

ing process beforehand, or *ex ante*, is basically that a choice has to be made – a *decision* must be taken in order to have something to assess in the first place. In doing so, presumably, we could not say anything about what underlying preference, considerations, or motivation will lead the agent to her final decision, because we would be unable to make sure that we have enumerated all the underlying motivational considerations of the choice set. Such an approach would be *descriptively* satisfying only if the agent was merely influenced by the bare observable outcome and commonly known information thereof[11]. However, this is unlikely to be the case in situations which interest us here: we would not be able to consider the possibility of c^* in Jane's example even if we realized later on that she was in possession of the particular information that turned c into c^*. To follow a strict and rigid normative *ex ante* requirement in Jane's case – that is, to focus merely on the eventual decision and hence not differentiating between c and c^* – would be descriptively deeply unsatisfactory. However, it is not clear why allowing c^* would be so problematic on a normative-deductive level. Could there not be a middle-ground between the objectives?

In fact, Anand himself seems to imply an interconnection between the objectives. For him, axioms in themselves are empty (Anand, 1990: 98). They need to be applied to a domain via defined relations. Indeed, how and where axioms like transitivity are supposed to apply is the crux of the matter. On this basis, primitive construction in itself (and re-description thereof) is not so much the issue at stake; it might even be a rather useful tool in testing, for instance, outcomes of experiments (Anand, 1990: 98-99). According to Anand, we need to make sure that the primitives have 'substantive meaning' and a non-empty behavioral content (Anand, 1990: 99). This is how we can define and clarify the primitives and their relations to the domain. So, how do we determine the relevant information in the choice set of economic agents in order to obtain primitives with 'substantive meaning'?

To do this, we simply need to specify and describe a choice primitive such that it holds for the specific decision-making process of the agent. The description of the choice primitive has to correspond to the grounds on which the agent makes her decision and has thus to encompass the relevant motivational basis of the decision maker. By motivational I mean the set of considerations, beliefs and thoughts that *make* the agent take a decision. It is the set of factors that lead her to the conviction of acting on reasoned grounds. It is in virtue of this motivational basis – and with regard to this basis only – that an observer can legitimately assess the rationality of an action. Those reasons make the decision rational or not. An act is not rational by chance or because certain unconsidered or unknown factors *happen* to make an act or choice appear rational. A choice is not rational if it is identical to what a rational person would have chosen. It is rational if the 'right' and rational reasons have led the person to make that decision[12].

Consequently, we need to get the description right with regard to the agent's perception of the choice – this is the criterion that saves the description from being arbitrary. I admit that we hereby jettison a strictly 'hands-off' requirement because the underlying motivation might not be directly observable and thus not defined *ex ante*. However, I am not saying that the requirement is useless *per se*. In fact, the previously mentioned legitimate basis for assessing the rationality of agents itself represents a hands-off requirement. Basing this assessment only on the motivational grounds of an agent is less strict and more encompassing. *Pace* Anand, this focus on motivational grounds is not necessarily problematic, as we shall see now.

3.3 The impact on decision theory

Allowing for a re-description of the choice primitive after the fact need not be a problem so long as we base the re-description of the choice primitives on evidence external to the decision theorist, the experimenter herself, or the observer in general. This is based on the assumption that a direct influence of the investigating subject on its object of study needs to be avoided or controlled for. But beyond that point, we do not need to restrict ourselves unnecessarily from finding an action's real motive. The grounds of action might not be the directly observable: in Jane's case, her apparent choice for c would have been considered irrational if we had not allowed for a broader scope, involving *her* underlying motivation. Indeed, our primary focus is what leads the agent to her decision and not so much

what this ultimate choice is. Whatever the grounds for the decision are, these grounds need to serve as the final answer to the question whether a choice has been rational or not. Rationality is attributed to the underlying decision-making process of the agent; her choice is merely the output and result of this process.

Our aim in this paper has been to show that there is a possibility to preserve the axiomatic method and transitivity despite the challenge of experimental results. Let me now briefly address a major charge against this view. To describe the choice primitives in terms of the motivational reasons or beliefs of the agent might seem quite natural. But the former are not necessarily directly observable. This not only poses a problem concerning the proper observation and categorization of these motivations but also has problematic consequences for the demands on the observer or experimenter.

Whereas the first issue can be dealt with technically (by refining methods such as surveys or experiments, or finding appropriate proxies, etc.), the second issue goes deeper. It aims at current problems like data mining, that is, the selective interpretation and treatment of data to confirm a specific preconceived conclusion. Is it possible for the experimenter to not at all influence the social experiment? Admittedly, our benchmark is less demanding than what is required by current standards in rational choice theory (that is, strict 'hands-off'). However, we have seen in the previous sections that a benchmark of objectivity that satisfies the 'hands-off' criterion perfectly is rather unrealistic. Involvement of the experimenter is unavoidable, as can be seen in issues of, for instance, framing experiments[13]. It is a matter of degree, and one may criticize our account for giving too much leeway to this involvement. The complete absence of involvement desired by those critics, however, should be given up, since it is an illusion.

Let us return to Jane in order to consider some preconditions under which our account of re-description is likely to be successful. What if Jane reconstructs or invents a rational choice after the fact? She might be intelligent enough to make up a story that makes her *prima facie* intransitive decision appear rational after all. Here, the decision theorist will have to roll up her sleeves, soil her hands, and conduct an investigative case-by-case study asking whether Jane's new story makes sense. Furthermore, the experimenter would have to take a clear stand. Firstly, by openly defining and *defending* a normative benchmark of rationality by which behavior is assessed. Secondly, by eliciting a transparent criterion for the appropriate translation or (re-)description of the choice.

By loosening up the 'hands-off' requirement and allowing the experimenter to be involved we can still address issues of obvious contradictions in preferences. For alluding to the motivational grounds of an action does not bar us from assessing such cases on a normative level. Our benchmark of rationality and transitivity (or whatever other axioms we posit) still applies. If it turns out that Jane simply enjoys violating axioms, and, moreover, our c^* has never been the case or was deliberately neglected by Jane – then there is no reason not to frame her ultimate decision as irrational according to the benchmark. Evidently, at the end of the day it all depends on our definition of rationality. This, in turn, depends on the axioms we endorse beforehand to determine what we mean by rationality. Questioning an overly standardized understanding and application of rationality or doubting its underlying axioms, as done by Anand and others, is a desirable endeavor. However, we need not throw out the baby with the bath-water and by dismissing transitivity altogether. Nor do we have to chuckle less about Jane's anecdote.

4. Conclusion

In this paper, we looked at the puzzle of transitivity focusing on the example of Jane: is her intransitive choice to take the cheese cake necessarily irrational? We suggested a re-description of her choice according to motivationally salient considerations as a way to preserve transitivity. We argued against Anand's claim that any re-description is ultimately arbitrary, by showing that observable choice is not the sole criterion for describing preferences. Furthermore, we suggested that Anand's argument is based on an overly strict distinction between descriptive and normative decision theory. In our view, both approaches are interrelated. On such an account, we are no longer bound to a strong and fruitless dichotomy between questioning intuitive axioms and accepting irrationality on a large scale.

Acknowledgments

Many thanks go to Conrad Heilmann for his help and readiness to provide extensive feedback on several versions. Comments from Willem van der Deijl as well as two anonymous referees from ESJP have also contributed to substantial improvements of this essay.

Darian Heim (1987) is currently pursuing a PhD program at the Universitat Pompeu Fabra in Barcelona, where he works on theories of justice and the rights of immigrants and multiculturalism. He holds a research MA degree from the Erasmus University in 'Philosophy and Economics' and a BA from the University of Fribourg (Switzerland) with a major in philosophy and a minor in economics.

'Getting the Description Right' was written for the master's course 'Decision and Game Theory' with dr. H.C.K. Heilmann.

Notes

1. This anecdote is attributed to Sidney Morgenbesser as a case of violating the independence of irrelevant alternatives. We shall presently adapt and use the anecdote for our purposes of discussing transitivity.

2. Dietrich & List (2009) model an agent's preferences as influenced by 'motivational salient dimensions' of options. Evidently, the present approach is inspired by their account but not necessarily perfectly congruent with it.

3. Anand (2009) endorses the validity of the claims in his 1990 article.

4. All following, not further documented references relate to Anand (1990).

5. The contextualization in the present paragraph is mine and not literally found in Anand (1990).

6. I adapted Allais' findings to our example and neglected framing issues for the present purpose.

7. Kahneman & Tversky (1979) and Loomes & Sugden (1982), most seminally, extended on this 'Allais paradox'. Both of these stand in a tradition of gradually eroding the importance of transitivity in economic modelling.

8. Although we do not deal with them, accounts of rational choice that relativize or waive transitivity are acknowledged (cf. Anand (2009) and Hansson & Grüne-Yanoff (2011) for a representative list).

9. Although Anand argues against this conception, it underlies my own argument in particular in section 3.2.

10. Cf. Mandler (2005) for a detailed account of the nature of this relationship.

11. This implicit assumption of Anand is clearly inspired by the broadly shared behavioral framework among economists involving 'revealed preferences': whatever the agent chooses is what she de facto prefers. Our alternative and contrastive account, in turn, is inspired by Davidson (1974) or Sen (1977) and focuses on internal states as determinants of preferences.

12. Evidently, such an account involves several controversial philosophical premises – epistemological internalism or world-mind dualism, to provide some labels. Although I will not be able to deal with these considerations here, my aim is reached if I can show that there are argumentations that avoid Anand's dismissal of transitivity.

13. The matter her is, in a nutshell, that the way choices are described, presented and framed by the theorist has an influence on how people act despite the fact that the objective probabilistic outcomes are identical. This is analogous to Allais' (1953) or Kahneman and Tversky's (1979) findings.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. In: *Econometrica*, 21 (4), 503-546.

Anand, P. (1990). Interpreting Axiomatic (Decision) Theory. In: *Annals of Operations Research*, 23, 91-101.

Anand, P. (2009). Rationality and Intransitive Preference: Foundations for the Modern View. In: Anand, P., Pattanaik, P.K., & Puppe, C. (eds.) *The Handbook of Rational and Social Choice. An Overview of New Foundations and Applications*. Oxford University Press, Oxford, 156-172.

Davidson, D. (1974). Psychology as Philosophy. In: Davidson, D. (2001) *Essays on Actions and Events*, 2nd ed., Oxford University Press, Oxford, 79-89.

Dietrich, F. & List, C. (2009). A Model of Non-Informational Preference Change. In: *The LSE Choice Group Working Paper Series*, 5 (1).

Fishburn, P.C. (1991). Decision Theory: The Next 100 Years? In: *The Economic Journal*, 101 (404), 27-32.

Hansson, S.O. & Grüne-Yanoff, T. (2011). Preferences. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/preferences/. Last access: 21/12/2012.

Kahneman, D. & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. In: *Econometrica*, 47 (2), 263-292.

Loomes, G. & Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. In: *The Economic Journal*, 92 (368), 805-824.

Mandler, M. (2005). Incomplete preferences and rational intransitivity of choice. In: *Games and Economic Behavior*, 50, 255-277.

Sen, A. (1977). Rational fools: A Critique of the Behavioural Foundations of Economic Theory. In: *Philosophy and Public Affairs*, 6 (4), 317-44.