Incomplete preferences in disaster risk management

Nicolas Espinoza*

Department of Languages and Culture
Luleå University of Technology
97187 Luleå, Sweden
E-mail: nicolas.espinoza@ltu.se
*Corresponding author

Martin Peterson

Department of History and Philosophy of Science
University of Cambridge
Free School Lane, Cambridge, CB2 3RH, UK
E-mail: mbp24@cam.ac.uk

Abstract: This paper addresses the phenomenon of incomplete preferences in disaster risk management. If an agent finds two options to be incomparable and thus has an incomplete preference ordering, *i.e.*, neither prefers one option over the other nor finds them equally as good, it is not possible for the agent to perform a value tradeoff, necessary for an informed decision, between these two options. In this paper we suggest a way to model incomplete preference orderings by means of probabilistic preferences, and how to reveal an agent's incomplete preference ordering within a behaviorist framework.

Keywords: incomplete preferences; probabilistic preferences; incomparability; disaster risk; decision; choice; revealed preference.

Reference to this paper should be made as follows: Espinoza, N. and Peterson, M. (2008) 'Incomplete preferences in disaster risk management', *Int. J. Technology, Policy and Management*, Vol. 8, No. 4, pp.341–358.

Biographical notes: Nicolas Espinoza is a PhD student of Philosophy at the Luleå University of Technology in Sweden. His main area of research is value theory and its application to risk evaluation.

Martin Peterson is a Research Fellow at the Department of History and Philosophy of Science, the University of Cambridge, UK. His main areas of interest are the philosophy of risk, decision theory and ethics.

1 Introduction

A comprehensive disaster risk management process consists of the pre-disaster phase and the post-disaster phase. The pre-disaster phase includes risk identification, risk mitigation, risk transfer and preparedness; the post-disaster phase involves emergency response and rehabilitation and reconstruction. With a limited budget for this entire process, and subsequently for each individual objective of the process, a decision-maker (which may be an individual or a collective) must decide which objective should receive the most resources, and also, which of the attributes that satisfy the objective should be prioritised. However, there is the problem that:

> "there are circumstances where we can enhance the achievement of one objective only if we accept a degradation in the achievement of another objective. The judgment made about how much we are willing to give up in terms of achieving one objective to better achieve a second objective by a specific amount is referred to as value tradeoff. Value tradeoffs are difficult to make, but necessary as they are at the heart of making informed choices." (Keeney, 2007, p.588)

At the most general level a decision-maker faces a tradeoff between the pre-disaster phase and the post-disaster phase; for instance, money spent on mitigation now is money that cannot be spent on reconstruction later on. More specifically, consider for example the rehabilitation and reconstruction of New Orleans after hurricane Katrina. One way of reasoning is that since natural events become disasters only because populations inhabit areas not sufficiently prepared to withstand anticipated natural hazard events, simply relocating the population to a less vulnerable area is cheaper than rebuilding in the same place *and* protecting it. Although relocation is perhaps the cheaper alternative, at least in monetary terms, it forfeits hundreds of years of aggregated cultural value tied up in the present location of New Orleans. How does one, then, weigh the cultural 'benefits' of maintaining New Orleans and the costs of rebuilding hurricane protection? In a case like this, where fundamentally different values are at stake, there does not seem to be a conversion rate between the involved objectives.

In this paper we will discuss these kinds of cases, when tradeoffs are not only difficult but seemingly impossible. The main characteristic of such cases is that there does not exist a single measure, or scale, along which two objectives can be weighed, *i.e.*, the objectives are incomparable, and there thus seems to be no way of making an informed tradeoff between them.

The notions of 'value tradeoff' and 'incomparability' will be modelled in terms of the the decision-maker's (the agent's) preferences: There exists a value tradeoff between objectives O and P if and only if there are some degrees of achievement m and n such that the agent is indifferent between O to the degree m (the *option Om)* and P to the degree n (the option *Pn).* If it is false that the agent prefers one option to the other, and if it is false that the agent is indifferent, it follows that the options are *incomparable* and that the agent's preference ordering is *incomplete.*

Our aim is to first elaborate on what incomparability is and how it may be represented through incomplete preference orderings, and second, demonstrate what kind of choice behaviour one should expect from someone with incomplete preferences. These goals will be achieved by means of a *probabilistic* analysis of incomparability. Briefly put, advocates of the probabilistic theory maintain that preferences come in degrees, *i.e.*, that having a preference is not an all or nothing affair. It is helpful to illustrate this basic theoretical idea by using very simple examples not directly related to hurricane Katrina. Imagine, for example, that you are attempting to choose between a Porsche or a Jaguar. Introspection reveals that you have no settled opinion about which car you prefer the most (they are incomparable). The Porsche is faster and more reliable, but the Jaguar is sleeker and more elegant. To some degree, you prefer the Porsche to the Jaguar *and* the Jaguar to the Porsche at the same time, since you value all these attributes. This is not necessarily irrational. If you were to contact your local Jaguar dealer, you would think that the Porsche is better, but had you contacted the Porsche dealer, you would have thought that the Jaguar was really what you wanted to buy. Your incomplete preference ordering is not due to any lack of information about the cars. Your preference ordering would remain incomplete, even had you known all facts about the two cars. We argue that such indeterminate preferences can be analysed in terms of subjective probabilities for acts.[1] To say that the Porsche and the Jaguar are incomparable roughly means that the agent will choose the Porsche with some probability $0 < p < 1$, if given the opportunity. Obviously, this is not equivalent to saying that the *strength* of a preference may vary. That claim is trivial. The thesis defended herein is that even if one were to keep the strength of ones preference fixed, one may still prefer the Porsche to the Jaguar to degree p, and the Jaguar to the Porsche to degree $1 - p$, all things considered.

The remaining parts of the paper fall into five sections. In Section 2, we give a background where we present the basic notions of incomparability, and the most prominent incomparabilist argument, the small-improvement argument. We argue that four binary preference relations, the usual three plus incomparability, do not constitute a satisfactory analysis of rational preference. The main claim here is that it is not enough to simply generalise a theory to incorporate incompleteness, although it is the first necessary step; one must also accommodate the (infinitely) many various degrees of incomparability that are possible once incompleteness has been acknowledged. This is facilitated by the probabilistic interpretation of preference, presented in Section 3. We also show how the probabilistic interpretation of incomparability is utilised to reveal that an individual has incomplete preferences. In Section 4, we compare our theory of incomparability with previous accounts. By acknowledging that incomparability is a matter of degree, our analysis differs from previous analyses, *e.g.*, Danan (2003), which also aims at demonstrating how incomparability may be revealed through choice behaviour. In Section 5, we discuss how the theory of probabilistic preferences bears on decisions in disaster risk management. Finally, in Section 6, we discuss some potential problems with our analysis and indicate some possible solutions.

2 Previous accounts of incomparability

Jet fuel to Arlanda Airport in Stockholm, Sweden is delivered from the harbor to the airport in tank trucks. On their way to the airport the trucks must pass through central Stockholm. A risk analysis has shown that the probability for an accident in which a large number of people in the city centre are killed by an exploding tank truck is about 1 in 10.000 per year. Another alternative is to use an existing railway. If that alternative is chosen the cost will be considerably higher, but the risk that many people are killed in a catastrophe is avoided. There is, however, a risk that a nearby groundwater reservoir will be polluted if two trains collide. The probability for this event is about 2x10E-6 per year according to the best estimates available. Now, one might ask whether the choice between the two alternatives described above can really be made by assigning monetary values to the different kinds of accidents and then calculate the expected monetary loss from each alternative, and finally compare it to the economic gains of implementing the two alternatives. Arguably, it can be questioned whether it is really meaningful to compare the risk that many people die to the risk that a ground water reservoir is polluted. Anyone wishing to account for human choice behaviour in such cases has to take people's unwillingness to make those kinds of comparisons into account. There is no *a priori* reason to exclude that options are sometimes incomparable.

However, for the social scientist who accepts Revealed Preference Theory (RPT) there is no conceptual space for incomplete preferences. Unless choices violate the so called Weak Axiom of Revealed Preference (WARP), they automatically reveal a preference. As will be explained in subsequent sections, WARP implies that an agent either prefers x to y, or y to x, or is indifferent between them. There are no more preference relations one can have, because all claims about preferences must be based on observations of pair-wise choices. (Incidentally RPT cannot distinguish indifference from strict preference unless a series of choices are observed.) Therefore, given that incomplete preferences do exist, we must either abandon RPT, as is suggested by Hausman (2000), or alter the theory somehow to accommodate incomparability.[2] The second alternative, which will be explored here, simply amounts to recasting the notion of preference into one which can couch talk of incomparability and which is also compatible with the basic tenet of RPT, that is, the claim that actual choice behaviour is the only evidence we have for justifying statements about preferences.

The notion of preference inherent in this theory is not compatible with RPT in a strict sense, nor shall we try to make it so. This paper merely seeks to achieve a rather modest aim, namely to present a new concept of incomparability, that is consistent with the claim that preference relations can be revealed through choice behaviour.

The probabilistic theory, if correct, will yield more accurate predictions than RPT. An advocate of RPT observing a consumer who chooses x over y seven times out of ten cannot use this information for making predictions about future choices, because the only way to rationalise such behaviour is to assume that the agent's preference changed while choice data was collected. However, if we accept the probabilistic theory and observe someone who chooses x over y seven times out of ten, we could use this evidence for predicting that the agent will choose x next time with a probability of about 0.7.[3]

Instead of confining the analysis of preference to three preference relations – preference, indifference, and dispreference – philosophers and (some) social scientists nowadays maintain that exactly one of *four* relations holds between a pair of mutually exclusive and exhaustive objects. These relations can be defined in terms of a single primitive preference relation, '*at least as preferred as*'. Let the pair $\langle O, \succeq \rangle$ be a comparison structure in which O is a set of objects $\{x, y, \ldots\}$ and $\succeq$ is a binary relation on O such that $x \succeq y$ if and only if x is at least as preferred as y. Then:

1 the agent strictly prefers x to y iff $(x \succeq y) \wedge \neg(y \succeq x)$

2 the agent strictly disprefers x to y iff $\neg(x \succeq y) \wedge (y \succeq x)$

3 the agent is indifferent between x and y iff $(x \succeq y) \wedge (y \succeq x)$

4 the agent considers x and y to be incomparable iff $\neg(x \succeq y) \wedge \neg(y \succeq x)$.

Any theorist wishing to link preferences to choice behaviour by adopting definitions (1)–(4) will run into difficulties. First, it will be hard to distinguish strict preference from indifference. If it is observed that x is chosen in a pair-wise choice, an external observer cannot tell if x was chosen *because* x and y were considered to be indifferent, or because x was strictly preferred to y. Furthermore, it is hard to see what kind of behaviour could correspond to (4). If neither *x* nor *y* is chosen, *e.g.*, if no choice at all is made, it is after all not a choice between two mutually exclusive and exhaustive options. In fact, this is the reason why RPT does not acknowledge (4) as a possible preference relation. Whatever option the agent chooses, an advocate of RPT will argue that she (weakly) preferred that option.

Nevertheless, people often *say* and *feel* that certain options are incomparable. For instance, there is evidence of so-called sacred or protected values, which can be defined as "any value that a moral community implicitly or explicitly treats as possessing infinite or transcendental significance that precludes comparisons, trade-offs, or indeed any other mingling with bounded or secular values" (Tetlock *et al.*, 2000). Indeed, a growing body of empirical evidence suggests that not only do people express moral outrage when asked to perform trade-offs between a protected value and money, but they fail to be consistent in doing so (Tetlock, 2003). As a result, several researchers have noted that protected values cause problems for quantitative elicitation of values, as is done in cost-benefit analysis or decision analysis (Baron, 1997; Bazerman *et al.*, 1999). For instance, people who hold a protected value for forests might say that it is just as bad that a large forest is destroyed in a fire as a small one.

For a more detailed illustration of the effects of incomparability, recall the example with how to transport jet-fuel to Arlanda Airport. After seriously considering the advantages and disadvantages of numerous different alternatives, such as using an existing railway that will jeopardise a ground water reservoir in case of an accident, or using tank trucks that may kill dozens of people if they explode in the city centre, it may very well hold true that the decision-maker has no definitive opinion about which option is best. It may be very hard for the decision-maker to resolve which of these factors – clean environment or human health – matters most, and also how much of one factor one is willing to give up for the sake of the other. Given some considerations one may prefer the railway to the tank trucks, but also, given other considerations, one may prefer the tank trucks to the railway. However, all things considered one need not (fully) prefer one option to the other.

In a case like this it might seem natural to assume that since you do not prefer one to the other, it simply has to be the case that you are indifferent between them. This conclusion is, however, unjustified. Savage (1954) noted:

> "if the person really does regard f and g [two options] as equivalent, that is, if he is indifferent between them, then, if f or g were modified by attaching an arbitrarily small bonus to its consequences in every state, the person's decision would presumably be for whichever act was thus modified."

So, if one were indifferent, then a small improvement on either option (say a \$100 discount on the railway option) would tip the scale in favour of the improved option. We may safely assume, however, that such a small discount on an investment in new infrastructure costing a fortune would not make the decision-maker prefer the improved option.

The *small improvement argument* utilises this 'indifference test' to establish incomparability. Simply put, the argument claims that if one can establish that there is no preference for either alternative and then add a small bonus to one of the options, but still do not find that the agent prefers one to the other, this proves that the agent was not indifferent. The alternatives are thus shown to have been incomparable.[4] The problem is, of course, how to establish (by only observing choices) that there is no preference for either alternative.

Some people might feel that there is an obvious way in which one could establish that there is no preference for either alternative, and thereby use the 'indifference test' for linking incomparability to choice behaviour: If the agent is first offered to choose repeatedly between option *x* and *y,* and then between *x* and a slightly discounted version of *y*, and chooses each object five times out of ten, it might be reasonable to say that the two options must have been incomparable. This proposal is, however, problematic, for the following three reasons.

First of all, an external observer can never be sure that the agents tastes and beliefs have remained constant during the course of the experiment, not even if she is locked up in a laboratory. Secondly, the proposed procedure only works for repeated choices. In real life most decisions on investment in new infrastructure is taken just once, not ten – at least within a restricted period of time. So on this proposal, one will never be able to detect incomparability, or behaviour corresponding to incomparability, by observing market transactions. Thirdly, and this is the most important point, a definition of incomparability based on the 'indifference test' will run into severe difficulties if it turns out that the agent chooses option *x* over *y* not five, but say eight, times out of ten and option *x* over the discounted version of *y* seven times out of ten. This behaviour need not be irrational, but it is neither compatible with the notion of preference adopted in RPT, nor with the notion of incomparability considered here.

It should be noted that incomplete preference orderings can occur not only in multi-attribute choice situations in which there is no weighting function, *i.e.*, no trade-off, among the various attributes that contribute to the value of the choice option. Incomplete preferences can occur even in decisions involving only a single attribute. The beauty of Mozart's music can be judged to be incomparable to the beauty of Van Gogh's paintings, for instance.[5]

If we accept that incomparability is a real possibility, difficulties arise for theories in which the possibility of an incomplete preference ordering has earlier not been acknowledged. For a well known example, one of the axioms of von Neumann-Morgenstern utility theory is the completeness axiom (or comparability assumption). It assumes that any set of individual preferences is complete, *i.e.*, for any two mutually exclusive choice options *u* and *v,* exactly one of the following relations holds: $u \succ v$, $u \prec v$, or $u \sim v$. But, if incomparability exists then we do not have a system of complete preferences, and subsequently utilities cannot be assigned. Of course, von Neumann and Morgenstern were aware of the fact that "one may doubt whether a person can always decide which of two alternatives ... he prefers" (von Neumann and Morgenstern, 1947, p.28). They suggested that an amendment to this problem would be to formulate what they called a "many-dimensional vector concept of utility", but this was deemed "more complicated and less satisfactory" (p.29). However, since *The Theory of Games and Economic Behaviour* there have been several works targeted at generalising the theory to render it independent of the completeness assumption. Aumann (1962) was to pioneer this endeavour followed by Bewley (1986), and more recently Dubra *et al.* (2004). The effort to formulate expected utility without assuming a complete preference ordering entails generalising the theory to incorporate incomparability as a fourth preference relation. This paper adds to this ongoing effort by arguing that incomparability actually comprises a whole range of (probabilistic) preference relations.

If this claim is correct, the new versions of expected utility theory will not differentiate among a whole class of possible preferences, and will thus merely constitute a partial analysis.

Revealed preference theory is, of course, conceptually independent from expected utility theory. In RPT, preferences are identified through actual choice behaviour, so if x is chosen as y is available, then x is preferred to y. Furthermore, according to the *weak axiom of revealed preference*, if a person reveals a preference for x over y, then he must not also reveal a preference for y over x (the assumption of course being that the agent's preferences have not changed between the choice situations). Or, put in Samuelson's words, "if an individual selects batch one over batch two, he does not at the same time select batch two over one" (Samuelson, 1938, p.65). It should be noted, however, that 'reveal' may not be the most appropriate term to use in this context. For if we distinguish between what is sometimes referred to as psychological preferences, *i.e.*, agents' introspective tastes and beliefs, and behavioural preferences, *i.e.*, agents' observable choice behaviour, it is not the case that our behaviour necessarily reveals our preferences (Sen, 1973). We should rather interpret the axiom as saying that if a person chooses *as if* he prefers x to y, then if he is rational he will not choose as if he prefers y to x. But what if the agent is indecisive and does not have a preference for either option, nor is indifferent between them? RPT will nevertheless rationalise choice behaviour as if there was a preference in place.

Furthermore, in only allowing two behaviourally observable modes of preference (choice or no choice), RPT impoverishes the potentially infinite preferential landscape. The problem is that incomparability is analysed entirely in terms of negations; *i.e.*, as $\neg(x \succeq y) \wedge \neg(y \succeq x)$. However, there can be *preferential* differences between individuals that find the same options incomparable in that sense. For instance, someone who finds disaster relief option A and disaster relief option B *incomparable*, in terms of reducing human suffering, may ultimately *choose* to implement A rather than B six times out of ten, while someone else who also finds them incomparable may choose A nine times out of ten. Both individuals have incomplete preferences but yet their preferences are different somehow. The fact that the first agent is less likely to choose A (six is less than nine) shows that her preferential attitude is not *exactly* the same as that of the second agent. There is a difference here, which is of interest to anyone wishing to predict and explain human behaviour. But this difference cannot be coherently described in traditional analyses of preference. The difference has to do with the degree to which the two decision-makers feel indecisive, *i.e.*, the degree to which they consider the two items to be incomparable. Hence, what is needed is a more nuanced analysis of incomparability, one that is able to represent varying degrees of incomparability. As will be shown in the next section, the probabilistic analysis we propose can meet this challenge.

3 The probabilistic analysis

The probabilistic analysis of preference we propose is the following. Let $\langle \succ_p, O \rangle$ be a comparison structure in which O is a set of options $\{ x, y, \ldots \}$ and $\succ_p$ is a relation on O such that $x \succ_p y$ if and only if the subjective probability is p that x will be chosen over y. (We will return to the interpretation of subjective probability in the next section.) Also, let x^+ be a slightly improved version of x such that $x^+ \succ_1 x$.[6] Then:

1 the agent strictly prefers x to y iff $x \succ_1 y$

2 the agent strictly disprefers x to y iff $x \succ_0 y$

3 the agent is indifferent between x and y iff $(x \succ_{0.5} y) \wedge (x^+ \succ_1 y) \wedge (y^+ \succ_1 x)$

4 the agent considers x and y to be incomparable iff $(x \succ_p y) \wedge (x^+ \succ_{p'} y)$, where $0 < p < p' < 1$.

The intuition articulated in (1) and (2) is that x is strictly preferred to y just in case it is certain that the agent will choose x rather than y (or vice versa) if given the opportunity. According to the probabilistic notion of indifference articulated in Condition 3, the agent is indifferent whenever there is a 50-50 chance for each option to be chosen and a small improvement of one of the items would raise the probability from 0.5 to 1 that the improved item is chosen. Hence, if you are indifferent between Fanta and Pepsi at current prices, but get the opportunity to buy a Pepsi for one penny less, the probability that you choose Pepsi will rise to one; otherwise you were not indifferent. This notion of indifference can be contrasted with the following asymmetrical indifference relation.

(3*) The agent is asymmetrically indifferent between x and y iff

$$(x \succ_p y) \wedge (x^+ \succ_1 y), \text{ where } 0 < p < 0.5 \text{ or } 0.5 < p < 1.$$

If the probability is, say, 0.7 that you will choose x over y, but it is certain that you will choose x^+ over y, then you are asymmetrically indifferent between the two objects in the sense specified by (3*). We by no means claim that this is a common type of probabilistic preferences, but it is worth mentioning as a genuine possibility – it cannot be excluded *a priori* that people sometimes behave like this.

Incomparability is modelled by (4). The basic idea is that the agent considers two objects to be incomparable in case the probability that x will be chosen over y is higher than zero but less than one, given that a small improvement will increase the likelihood that the improved object is chosen a just little bit. From a probabilistic point of view, the difference between indifference and incomparability is thus straightforward. An addition of a small bonus to x in case the agent is indifferent will increase the probability that the agent chooses x from 0.5 to 1; see (3) and (3*). However, if an equally small bonus is added to x when this object is incomparable to y, the probability that x is chosen would presumably increase just a little bit, say from 0.5 to 0.51. Hence, indifference is a probabilistic preference that is extremely sensitive to slight modifications of the alternatives, whereas incomparability is not.

Traditionally, a strict binary (non-probabilistic) preference is assumed to be irreflexive, asymmetric, and transitive. The corresponding properties of the probabilistic preference relation are as follows.

P-Reflexivity: $x \succ_{0.5} x$

P-Symmetry: If $x \succ_p y$, then $y \succ_{1-p} x$

P-Transitivity: If $x \succ_p y$, and $y \succ_q z$, then $x \succ_r z$, where $r = pq/[pq + (1-p)(1-q)]$.

The formula for P-Reflexivity may look somewhat arbitrary. However, note that any other value than 0.5 will imply a contradiction in conjunction with P-Symmetry, which we take to be a self-evident property. Arguably, it is best to think of the choices described by P-Reflexivity as between uniform types: P-reflexivity is not the claim that the agent will choose some individual object with probability 0.5 when compared with itself; the idea is rather that in a choice between one x and another x, the probability is 0.5 that the agent will choose the first x.

The formula for P-Transitivity is more complex. In order to justify this property Luce's well-known Choice Axiom has to be invoked. For this purpose, it is appropriate to adopt a slightly more complex notation: Let $p(x > B)$ be the probability that x is chosen from the set of alternatives B, and let $p(A > B)$ mean that when A is a subset of B the probability is p that the chosen alternative is an element in A. Assume that all probabilities are nonzero. Then, according to the Choice Axiom, the probability that x is chosen from B equals the probability of choosing x from A, where A is a subset of B, multiplied by the probability that the chosen alternative is an element in A. In symbols:

Choice Axiom If $A \subset B$, then $p(x > B) = p(x > A) \cdot p(A > B)$.

The Choice Axiom and the axioms of the probability calculus imply P-Transitivity. This was first proved by Luce (1959); we shall not repeat the proof here. It is beyond the scope of the present paper to offer any positive support or justification of the Choice Axiom. However, a merit of P-Transitivity is that the traditional notion of indifference will remain a transitive relation, because $0.5 \cdot 0.5/[(0.5 \cdot 0.5) + (1 - 0.5)(1 - 0.5)] = 0.5$. Hence, if $x \succ_{0.5} y$, and $y \succ_{0.5} z$, then $x \succ_{0.5} z$. For the same reason, the traditional notion of strict preference is also a transitive relation, because $1 \cdot 1/[(1 \cdot 1) + (1 - 1)(1 - 1)] = 1$. Hence, if $x \succ_1 y$, and $y \succ_1 z$, then $x \succ_1 z$.

In order to illustrate what kind of assumption is at stake in the Choice Axiom, suppose that an agent is about to choose a bottle of wine from a list containing two red and two white ones. The Choice Axiom then tells us that it should not matter if the agent divides his choice into two stages, and first choose between red and white wine and then between the wines in the chosen subset, or choose directly which of the four wines to order. Hence, if the agent is indifferent between red and white wine in general, as well as between the two red wines and the two white ones, the probability that a particular bottle will be chosen is $1/2 \cdot 1/2 = 1/4$. There are, however, seemingly rational preferences that violate the Choice Axiom, as first pointed out by Debreu (1960). The following example was suggested by Coombs *et al.* (1970): At the restaurant, an agent is indifferent between seafood and meat, as well as between steak and roast beef. The menu comprises only three dishes: x lobster, y steak, and z roast beef. Let B be the entire menu, let A be the set comprising x and y, and let A' be the set of y and z. Now, given the notation introduced above, since you are indifferent between seafood and meat, $p(x \succ B) = \frac{1}{2}$. However, $p(A \succ B) = 1 - p(z \succ B) = 1 - p(z \succ A') \cdot p(A' \succ B) = 1 - \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$. Hence, the choice axiom implies that $p(x \succ B) = p(x \succ A) \cdot p(A \succ B) = \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{8}$.

In our view, advocates of the choice axiom should not feel too worried about this counter example. Several authors have suggested that the root of the problem lies in the individuation of alternatives.[7] Lobster, steak, and roast beef are not alternatives at the same level. Lobster belongs to the category seafood, and could equally well be replaced with tuna, or any other sea food dish. However, for Debreu's example to work, neither steak nor roast beef can be replaced with some other meat dish, say kebab, because then it would no longer be certain that the agent will remain indifferent between the two meat dishes. Arguably, the moral of Debreu's example is that alternatives must be individuated with care. More precisely, the Choice Axiom should be taken into account already when alternatives are individuated. It can thus be conceived of as a normative requirement for how alternatives ought to be individuated: They should be individuated such that the Choice Axiom holds. Arguably, this does not transform the Choice Axiom into a tautology. It rather imposes a normative constraint at a very basic level.

We are aware that this point about further individuation is not uncontroversial. Broome (1991, pp.98–100) discusses a similar defense of sure-thing principle in expected utility theory, and notes that many economists find the strategy somewhat arbitrary. The main thrust of the criticism is that the individuation of alternatives ought to be justified on independent grounds. However, we believe that Broome's method for responding to this criticism is essentially correct: "We must have *some* principle for [individuating alternatives]", and there is no reason to believe that such non-arbitrary principles could, at least in principle, be formulated (Broome, 1991, p.100).

Before closing this section, we wish to point out that Luce's reason for assigning probabilities to choices had nothing to do with our present concerns about incomparability. He assigned probabilities to choices in an attempt to address an empirical problem in psychology, *viz.* to explain why people sometimes behave in ways that contradict the prescriptions of traditional decision theory. Luce's explanation was that people sometimes fail to discriminate which alternative is best for them. More precisely, Luce explicitly claimed that all:

- alternatives faced by the agent have 'true' utilities, but
- sometimes the agent cannot fully perceive these utilities
- therefore occasionally chooses sub-optimal alternatives (see Luce, 1959).

The higher the probability is that an alternative is chosen, the better is the discrimination. This is parallel to a discrimination problem in which an agent cannot tell which of two stones is heaviest. There is a true fact of the matter, but due to limited perceptual capabilities, the truth cannot be revealed.

Judged from a contemporary perspective, it seems difficult to justify the assumption that people make probabilistic choices because there are utilities out there that they fail to perceive. The concept of utility is closely related to mental states, and with a few notable exceptions it is widely agreed that mental states are internal and readily accessible through introspection. Therefore, in order to avoid the kind of externalism proposed by Luce the present paper advocates an alternative, subjective interpretation of probabilistic choices. According to this interpretation, there are no external utilities that people fail to discriminate among. Probabilities are rather assigned to choices for indicating that there is a certain amount of genuine indeterminacy involved. Consequently, the expression $x \succ_{0.7} y$ does not mean that the agent tends to choose x over y seven times out of ten. It rather means that the agent's degree of belief that x will be chosen corresponds to the number 0.7.

4 Relation to previous theories

This section describes how the probabilistic theory is related to previous contributions.

Of course, the traditional algebraic representation of preference can only distinguish four preference relations, but as we have seen, four may not be enough. Our probabilistic theory amends this deficiency by acknowledging the possibility of more or less indeterminate preferences. However, it could of course be argued that it would be more fitting to say that such indeterminate preferences are *vague*, in the sense that it is neither true nor false that *x* is preferred to *y* or that *y* is preferred to *x.* Indeed, the intimate connection between incomparability and vagueness has been made plausible by Broome (1997).[8] By adopting vagueness-terminology it becomes natural to say that it is *truer* that *x* is preferred to *y* than it is true that *y* is preferred to *x*, even though it is not *entirely* true that *x* is preferred to *y; i.e.*, the indeterminateness between *x* and *y* can be seen as a matter of degree.

The claim that the agent's underlying mental state is vague or indeterminate may seem controversial. However, this should be understood in the same way as the mental state of for example hunger. When someone asks you if you are hungry it may be the case that you are hungry but at the same time that you are not hungry – your answer perhaps being "I'm not hungry, but I can eat!" It would be strange to press the point and demand of you to answer yes or no, since obviously hunger is a matter of degree. Accordingly, given a certain degree of hunger, you will eat some of the times and others you would not.

Although the notion of vagueness may lend itself to modelling psychological preferences, it is nevertheless the case that choices are distinct; there is no vagueness in observed choice behaviour. How then can preferential vagueness (indeterminacy or indecisiveness) and the degree of indeterminacy be conveyed through choice? As we have seen, advocates of the probabilistic view believe that indeterminacy can be modelled by ascribing subjective probabilities to choices. If an agent considers it highly probable that *x* will be chosen over *y* her preference is less indeterminate than if she considers both choices to be equally probable. Hence, a preference is not assumed to be a determinate mental state that motivates action, but rather a predictive hypothesis of future actions based on subjective probabilities (cf. Peterson, 2006). So, if someone says that they have a strict preference of *x* to *y*, this means that they assign the subjective probability of 1 to the event that they will choose *x* over *y.*

In recent years there has also been a few other proposals targeted at the problem of revealing incomparability through choice behaviour (*e.g.*, Danan, 2003; Eliaz and Ok, 2006). Common to most of them is that they are non-probabilistic, and therefore do not provide the means to differentiate among different degrees of incomparability. In the following we shall contrast our own theory to one of the most noteworthy non-probabilistic models, *viz.* the 'preference for flexibility' model proposed by Danan (2003).

Danan (2003) makes the distinction between behavioural and cognitive preferences. In his vocabulary, behavioural preferences refer to those that are revealed through observable choice behaviour under the revealed preference methodology: the agent prefers *x* to *y* if and only if the agent chooses *x* over *y.* Cognitive preferences, on the other hand, cannot be elicited by choice behaviour, unless they are identified with revealed preferences. The identification of cognitive preferences with behavioural ones would be a mistake however, given the possibility of, amongst other things, incomplete preferences.[9] Under the behavioural interpretation, incomplete preferences are ruled out by the fact that the agent is always forced to choose between x and y, as explained in Section 2. However, to solve the methodological problem of not being able to reveal cognitive preferences via choice behaviour, Danan designs a simple empirical test.

> "The key idea of the model consists in extending the set of alternatives to the power set of menus over these alternatives, and assuming that the agent is unable to compare x and x' if and only if she prefers the menu $\{x, x'\}$ to both x and x', *i.e.*, if she has a *preference for flexibility*." (Danan and Ziegelmeyer, 2004)

Of course, the plausibility of Danan's model depends crucially on the assumption that a preference for flexibility is really indicative of incomparability, *i.e.*, that an agent facing incomparable options should prefer to make the choice at a later time (at which he might hopefully have made up his mind). However, in our view, this assumption is troublesome. Firstly, this new choice situation does not escape the basic problem that the model was designed to resolve; we now face the problem of revealing the cognitive preference between three alternatives, x, y, and $\{x, y\}$, instead of just two. In this new case one cannot exclude the possibility that the agent has incomplete preferences among all three alternatives. Just as before, the fact that the agent now chooses the third alternative, the flexibility option, does not imply that the agent cognitively prefers this option, since, again, it is a forced choice. Furthermore, there is no way of determining whether the agent is actually indifferent between the third option and one of the other two. Moreover, even if we disregard the previous point there is still the basic problem that you cannot deduce the preferential relation between two alternatives of a pair-wise choice by inferring some third alternative; it is then simply a different choice situation.

Of course, the probabilistic analysis we propose is not entirely new. The most important historic reference is Luce's (1959) monograph on individual choice behaviour, mentioned in Section 4. In order to better spell out how the present work relates to previous contributions originating from that monograph, it is helpful to invoke the distinction between two kinds of probabilistic choices suggested by Luce and Suppes (1965). They authors distinguish between *constant* utility models and *random* utility models. The difference between the two models can be demonstrated by separating the individual choice process into two steps: First, the decision-maker assesses the utility of each alternative, and second, she derives a choice based on the decision rule of utility maximisation. According to the constant utility model the first step of this process is deterministic while the second step is not, *i.e.*, choice is a probabilistic function of preferences or utilities (Luce, 1959). The random utility model (*e.g.*, Manski, 1977) states that the individual, in the second step, always chooses the alternative with the highest utility. However, the first step should be viewed as probabilistic. In both of these models preferences are regarded as subjective mental states, which are not available to direct observation. It is assumed that the probability measure, whether it refers to preference or choice, approximates a true underlying preference. The preference is thought to be directly available to the subject, but because the subject may be imprecise in assigning utilities, or there is a communicative discrepancy between the experimenter and the subject, a probabilistic model is needed to cancel out error.

We should be careful to note that our model of probabilistic preferences differs fundamentally from these older probabilistic models. Most importantly, in the present analysis probabilities: (i) are ascribed to choices and (ii) are subjective.

With regards to (i) we should distinguish our theory from other probabilistic theories. Ascribing a probability to a preference, rather than to a choice, implies a belief that some particular strict preference will be adopted at some time in the future. This is precisely what is maintained by the stochastic theory of random utility (Manski, 1977), but this should not be confused with the analysis proposed here. Instead probabilities are ascribed to choice. The ascription of probabilities to choice should, however, not be confused with the other stochastic theory, constant utility theory (Luce, 1959). According to that theory preferences are determinate but choices are random. As explained above, Luce explicitly assumed that all alternatives faced by the agent have 'true' utilities, but that the agent fails to discriminate between them and therefore occasionally chooses sub-optimal alternatives. In our analysis, there is no need to assume that there are utilities out there, which people fail to perceive.

With regards to (ii), as was already explained in Section 3, when we say that x is 0.7-preferred to y it does not mean that the agent chooses x over y seven times out of ten, but it does mean that the agent will have this subjective tendency. Hence, the agent should be willing to accept the corresponding bet the she at the present moment will choose x over y. For this reason, the probabilistic interpretation we propose also works for single choices situations, whereas previous probabilistic theories only work for repeated choices.

So, to finally recapitulate, under the view we propose incomparability is a preference relation best represented by the ascription of subjective probabilities to choice. This view differs significantly from vagueness view held by Broome (1997), the deterministic choice model of Danan (2003), as well as the probabilistic choice models of Luce (1959) and Manski (1977).

5 Motivating the probabilistic analysis

We started off by mentioning the tradeoff between the pre-disaster phase and the post-disaster phase, and the weighing of incomparable values in the aftermath of hurricane Katrina. In order to illustrate the practical importance of our analysis of incomparability, it may be valuable to point out exactly what what goes missing in an analysis that does not take incomparability into account.

We need first ask the question, 'how *ought we to* prioritise and make decisions regarding disastrous risks?' One way to answer this question is to say that, obviously, decisions that concern disastrous risks must be at least as conscientious as decisions concerning non-disastrous risks. Indeed, the scale of potential immense human suffering and loss of material value in disastrous risks, places very high demands on the decision-maker. While a managerial mis-prioritisation in an ordinary risk management scenario will perhaps result in the needless suffering or death of one or two people, which is tragic enough, managerial errors in the disaster risk case may effect hundreds or even thousands. Thus, a good start is to consider what constitutes a managerial decision error in non-disastrous risk scenarios, and then impose even more stringent demands in the disaster risk case to avoid these errors.

Most decision-makers would surely agree that any risk decision ought to be *rational*, which in other words means that it should be at least:

- systematic
- consistent
- non-arbitrary.

That decision procedures be systematic is usually thought to be important because we want to avoid emotionally guided and unpredictable decisions. Most importantly, in order to determine if decision-makers made the right decision, given the context they were in at the time of the decision, we need to be able to follow the train of thought or method that they employed. Also, only if there is some system to the decision-making process is there any hope that it can improve over time. Without a systematic approach it is difficult to pinpoint what went wrong or what could be improved the next time a similar situation arises.

To spell out the consistency requirement we consider rational choice theory. A common view in the theory of rational choice is that our choices, to be rational, must correspond to our preferences. For instance, if an agent has the opportunity to choose between A and B and happens to prefer A over B then the agent is rationally required to choose A, rather than B. Furthermore, it is widely held that the preferences of a rational agent are subject to certain rationality constraints. For instance, if A is preferred over B then B cannot be preferred over A (asymmetry of preference), and if A is preferred over B, and B is preferred over C, then A must be preferred over C (transitivity of preference). Both of these aspects of rational choice, the correspondence between preference and choice, and the rationality constraints on preferences, are if not universally accepted, at least considered central to the theory of rational choice.

An important, but less discussed, aspect in rational choice is the non-arbitrariness condition. So far we have seen that an agent is considered rational if she makes systematic choices that reflect a consistent set of preferences that abide to a set of rationality constraints. However, even if these requirements are met, an agent could conceivably have quite arbitrary preferences. The preferences in themselves ought therefore to satisfy some minimal condition, namely that they reflect the value relationship of the alternatives under consideration. Roughly, if A is better than B the agent ought to prefer A over B. And, if A and B are equally as good the agent ought to be indifferent between them. Where things can go wrong, then, from the bottom and up, if we do not acknowledge incomparability, are the cases when the agent either has a strict preference between A and B even though A and B are incomparable, or chooses as if there is a strict preference in place even though she only has a partial one.

6 Two objections to the probabilistic analysis

This section addresses two objections to the probabilistic theory. Firstly, according to what has been claimed here, statements about preferences are purely cognitive statements.[10] When you say that you prefer something, this is a factual statement about how the world is (since you are talking about your own future behaviour), not about how the world *ought* to be. Hence, it seems that the evaluative aspect of a preference has been lost in the analysis. In reply to this objection, we would like to point out that a prediction about your own future behaviour is of course *indirectly* influenced by your emotions and desires; if you have a strong emotional aversion against spaghetti, you will probably predict that you will not choose spaghetti at the restaurant. So the evaluative aspect is in fact there; but since emotions cannot be *directly* observed, we have suggested a way in which they can be indirectly observed. This is in line with the main idea of revealed preference theory, which is, in a nutshell, that all we can say about preference is what we can say about choice, and that we should say nothing else; all we can do is rationalise choice behaviour *as if* there was a preference in place.

What about this – if anything – is unsatisfactory from an evaluative point of view? Obviously, there would be no problem if choices always corresponded exactly to our preferences, *i.e.*, if choices were, in a sense, proof of how we thought the world ought to be. The problem is, however, that we may sometimes prefer one thing but still choose another – perhaps by mistake, weakness of will, coercion, *etc.* – in which case the choice fails to be *normatively indicative*. However, despite that these kinds of non-indicating choices are possible, it is not unreasonable to think that they only constitute a small portion of all the choices we make. Most of the time we *do* choose what we prefer, and thus most of the time descriptions of what *is* the case adequately captures what *ought* to be the case.

A more serious problem, we think, is that many preferences will be corrupted, so to speak, by forced description as exactly one of strict preference, strict dispreference or indifference. We have tried to file away some of this crudity by generalising the notion of preference so that it can also accommodate varying degrees of incomparability and the respective corresponding behaviour. Hence, by providing a choice oriented theory that can accommodate *all* modes of valuation (better than, worse than, equally as good as, *and* incomparable to), we have reduced the portion of choices that fail to be normatively indicative.

A further objection that may be raised is the following. In the theories proposed by Ramsey (1926), DeFinetti (1974), and Savage (1954), probability is defined in terms of betting behaviour. Hence, on this view, if someone holds the preference $x \succ_{0.7} y$ he or she would be willing to accept a seven to ten bet that x will be chosen over y. Suppose for instance that John wishes to measure the subjective probability that he will buy a Jaguar rather than a Porsche. If John judges \$7 to be the highest price he is prepared to pay for a bet in which he wins \$10 if the proposition x = 'John buys a Jaguar' turns out to be true, but wins nothing otherwise, then his subjective probability for x is approximately 7/10, given that his utility of money is linear. For advocates of the Bayesian approach, it is important to distinguish between the case in which the agent assigns subjective probabilities to his *own* choices, and the case in which an external observer assigns subjective probabilities to someone else's choices. The second case does not give rise to any trouble; but the first has to be examined more closely: When the agent assigns subjective probabilities to his own choices, this readiness to accept bets becomes an instrument for measuring his own underlying (mental or non-mental) dispositions to act. For example, if he considers \$5 to be a fair price for a bet in which he wins \$15 if the proposition x = 'I will buy a Jaguar' turns out to be true, but wins nothing otherwise, his subjective probability for x is approximately 5/15, as explained above. However, as pointed out by Spohn (1977) and Levi (1989), the use of bets for measuring subjective probabilities for one's own acts give rise to the following problem:

- If an agent ascribes subjective probabilities to his present acts, then he must be prepared to take on bets on which act he will eventually choose.
- By taking on such bets, it becomes more attractive for the agent to perform the act he is betting on: It becomes tempting for the agent to win the bet by simply choosing the alternative that will make him win the bet.
- Therefore, the 'measurement instrument' (the bets used for eliciting subjective probabilities) will interfere with the entity being measured, that is, the agent's subjective degree of belief that he will choose a certain act.

Therefore, the agent's subjective probability will not reflect his preference.

However, even though an interesting argument, we believe that the measurement theoretical problem identified by Spohn and Levi can be avoided. Firstly, note that the qualitative theory defended in the previous section is not affected by this problem; this is in fact our main reason for not proposing a Bayesian interpretation of subjective probability. Secondly, we believe that Bayesians can control for this measurement theoretical effect in at least two ways. To start with, the betting situation could be modified such that the agent asks a well-informed friend who knows him very well to do the betting, without telling the agent whether he will win or loose the bet if the alternative he bets on is chosen. By separating the belief-generating mechanism from the deliberative mechanism in this manner, the measurement theoretical problem discovered by Spohn and Levi can no longer arise. For a detailed discussion of this proposal, see Rabinowicz (2002). Secondly, the stakes of the bets offered can be kept to a minimum, such that they become virtually negligible in comparison to the value of the outcomes. The point is that if we are aware of the problem, we can adjust the measurement process so that the disturbance is rendered minimal or non-existent. The measurement theoretical argument presupposes that net gain made by choosing in accordance with one's bets is of a certain size. We can block that assumption by setting a suitable limit for the bet that is relative to each decision situation. This point is explained in more detail in Peterson (2006).

Hence, even though it seems that Spohn and Levi are right that a measurement instrument based on betting dispositions may affect the outcome of the measurement process, the relevance of this observation should not be exaggerated. If we are aware of the problem, we can adjust the measurement process accordingly. The underlying phenomenon that is being measured – the subjective degree of belief – surely exists, no matter what the measurement theorist argues.

7 Conclusions

In this paper we have introduced the notion of probabilistic preferences in order to give an analysis of incomparability. While previous analyses of preferences, such as revealed preference theory, fail to take incomparability into account, the probabilistic analysis is able to represent varying degrees of incomparability that would otherwise be erroneously represented as strict preference or indifference. The theory thus enables a more nuanced analysis of risk decision behaviour, which is an advantage when considering decisions that preceed grave outcomes.

References

Aumann, R.J. (1962) 'Utility without the completeness axiom', *Econometrica*, Vol. 30, No. 3, pp.445–462.

Baron, J. (1997) 'Biases in the quantitative measurement of values for public decisions', *Psychological Bulletin*, Vol. 122, pp.72–88.

Bazerman, M.H., Moore, D.A. and Gillespie, J.J. (1999) 'The human mind as a barrier to wiser environmental agreements', *American Behavioral Scientist*, Vol. 42, pp.1277–1300.

Bewley, T. (1986) 'Knightian uncertainty theory: Part I', Cowles Foundation Discussion Paper No. 807.

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

Broome, J. (1997) *Is Incommensurability Vagueness?*, in Chang (1997).

Carlson, E. (2004) 'Broomes argument against value incomparability', *Utilitas*, Vol. 16, No. 2, pp.220–224.

Chang, R. (Ed.) (1997) *Incommensurability, Incomparability, and Practical Reason*, Harvard University Press.

Chang, R. (2002) 'The possibility of parity', *Ethics*, Vol. 112, pp.659–688.

Coombs, C.H., Dawes, R.M. and Tversky, A. (1970) *Mathematical Psychology: An Elementary Introduction*, Prentice-Hall.

Danan, E. (2003) 'A behavioural model of individual welfare', Mimeo, Universite de Paris 1.

Danan, E. and Ziegelmeyer, A. (2004) *Are Preferences Incomplete? An Experimental Study Using Flexible Choices*, Mimeo.

Debreu G. (1960) 'Review of R.D. Luce, individual choice behavior: a theoretical analysis', *American Economic Review*, Vol. 50, pp.186–188.

DeFinetti, B. (1974) *Theory of Probability: A Critical Introductory Treatment*, Wiley and Sons.

Dubra, J., Maccheroni, F. and Ok, E.A. (2004) 'Expected utility without the completeness axiom', *Journal of Economic Theory*, Vol. 115, pp.118–133.

Eliaz, K. and Ok, E. (2006) 'Indifference or indecisiveness? Choice-theoretic foundations of incomplete preferences', *Games and Economic Behaviour*, Vol. 56, pp.61–86.

Hausman, D.M. (2000) 'Revealed preference, belief, and game theory', *Economics and Philosophy*, Vol. 16, pp.99–115.

Keeney, R.L. (2007) 'Modeling values for anti-terrorism analysis', *Risk Analysis*, Vol. 27, pp.585–596.

Levi, I. (1989) 'Rationality, prediction, and autonomous choice', *Canadian Journal of Philosophy*, Vol. 19, pp.339–362, reprinted in Levi, I. (1997) *The Covenant of Reason*, Cambridge University Press.

Luce, R.D. (1959) *Individual Choice Behaviour: A Theoretical Analysis*, Wiley.

Luce, R.D. and Suppes, P. (1965) 'Preference, utility, and subjective probability', in R.D. Luce, B. Bush and E. Galanter (Eds.) *Handbook of Mathematical Psychology*, Wiley, Vol. 3, pp.249–410.

Manski, C.F. (1977) 'The structure of random utility models', *Theory and Decision*, Vol. 8, pp.229–254.

Peterson, M. (2006) 'Indeterminate preferences', *Philosophical Studies*, Vol. 130, pp.297–320.

Rabinowicz, W. (2002) 'Does practical deliberation crowd out self-prediction?', *Erkenntnis*, Vol. 57, pp.91–122.

Ramsey, F.P. (1926) *Truth and Probability*, in H.E. Kyburg and H.E. Smolker (1964).

Samuelson, P. (1938) 'A note on the pure theory of consumers', *Economica*, Vol. 5, pp.61–71.

Savage, L.J. (1954) *The Foundations of Statistics*, 2nd ed., Wiley and Sons.

Sen, A. (1973) 'Behaviour and the concept of preference', *Economica*, New series, Vol. 40, No. 159, pp.241–259.

Spohn, W. (1977) 'Where Luce and Krantz do really generalize savage's decision model', *Erkenntnis*, Vol. 11, pp.113–134.

Tetlock, H., *et al.* (2000) 'The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals', *Journal of Personality and Social Psychology*, Vol. 78, pp.853–870.

Tetlock, H. (2003) 'Thinking the unthinkable: sacred values and taboo cognitions', *Trends in Cognitive Sciences*, Vol. 7, pp.320–324.

Von Neumann, J. and Morgenstern, O. (1947) *Theory of Games and Economic Behaviour*, Princeton University Press.

Notes

1 This idea was recently proposed by Peterson (2006).

2 Our use of the term incomparability is one that is common in decision theory and philosophy of economics. We do not wish to confuse this notion of incompa-rability with other possible notions, which may perhaps be found within ethical and value theoretical discourses. Here, incomparability refers merely to the plausible situation where an agent neither has a strict preference, nor is indifferent, between two choice alternatives. We will not be exploring the connection between the concepts of preference and value in this paper.

3 As is explained in detail elsewhere, we believe that probabilities should be interpreted subjectively, rather than as frequencies. However, observed frequencies can nevertheless be used as evidence for claims about subjective probabilities.

4 For a complete version of the argument see Chang (2002).

5 For a good overview of the philosophical literature concerning incomparability see 'Introduction' in Chang (1997).

6 x^+ can be conceived of as the smallest improvement of x that affects the agent's likelihood to choose this object.

7 Luce, in personal communication and Peterson (2006).

8 Broome's argument has not gone unchallenged. For a recent criticism see Carlson (2004). Carlson's critique does not, however, object to the idea that the bet-terness relation may be indeterminate. In fact, he even provides his own examples of when it is indeterminate whether one item is better than another.

9 The words 'amongst other things' refer here to various possibilities why the agent does not choose what she really prefers. For instance the agent may be coerced or she may simply make a mistake.

10 This objection was raised by Wlodek Rabinowicz, in conversation.