Three Analyses of Sour Grapes*

Brian Hill

Groupe HEC
1 rue de la Libération
78350 Jouy-en-Josas
France
hill.brian@wanadoo.fr
www.hec.fr/hill

13th April 2007

Abstract

The phenomenon of adaptive preferences – sometimes also known under the name of *sour grapes* – has long caused a stir in Social Theory. Among logicians, notably those in the dynamic logic or belief revision traditions, the question of preference change has recently seen a surge of interest. However, although the former question seems an instance of the latter, the theories of preference change proposed to date do not seem to give a firm handle on adaptive preferences, and certainly not the sort of deeper understanding which one might like. In this paper, the precise problem posed by adaptive preferences, as seen from the point of view of a theoretician who intends to model or understand the phenomenon, will be clarified, and three models of the phenomenon will be presented and compared. The general intention of the article is to sound out some of the wider consequences of the phenomenon for the project of modelling and understanding the relationship between decisions taken in different situations. Difficulties which arise when several decisions and several situations are involved shall be discussed, and an approach to these difficulties shall be suggested. This approach places particular demands on would-be models of the sour grapes phenomenon; these demands will shed light on the adequacy of the models proposed.

Keywords: Adaptive preferences; preference change; belief change; decision theory; belief and utility elicitation; representation theorems.

JEL Codes: B49, D89.

*The author would like to thank Philippe Mongin for inspiring discussion and attentive reading of early versions of this text.

The phenomenon of adaptive preferences – sometimes also known under the name of *sour grapes* – has long caused a stir in Social Theory, mainly because of its importance in the debate over utilitarianism (Elster, 1983; Sen and Williams, 1982). Among logicians, notably those in the dynamic logic or belief revision traditions, the question of preference change has recently seen a surge of interest (van Benthem and Liu, 2007; Hansson, 1995). However, although the former question seems an instance of the latter, the theories of preference change proposed to date do not seem to give a firm handle on adaptive preferences, and certainly not the sort of deeper understanding which one might like. In this paper, the precise problem posed by adaptive preferences, as seen from the point of view of a theoretician who intends to model or understand the phenomenon, will be clarified, and three models of the phenomenon will be presented and compared. The general intention of the article is to sound out some of the wider consequences of the phenomenon for the project of modelling and understanding the relationship between decisions taken in different situations. Difficulties which arise when several decisions and several situations are involved shall be discussed, and an approach to these difficulties shall be suggested. This approach places particular demands on would-be models of the sour grapes phenomenon; these demands will shed light on the adequacy of the models proposed.

In the first section, we will briefly introduce the phenomenon of sour grapes, and general notions allowing us to single out some of its important properties. In the second section, three analyses of sour grapes shall be proposed, each based on a different intuition regarding the phenomenon. In the final section, we shall consider the important problems that the phenomenon poses for theories of beliefs, preferences and decisions, illustrating the issue with a comparative consideration of the three analyses.

1 What is the phenomenon of sour grapes?

In La Fontaine's fable, the fox approaches a tree, attempts to reach the grapes, and, realising that he cannot, turns away, saying to himself that they were sour.

This sort of phenomenon – and we shall be mainly considering this story in this paper – poses an interesting challenge to models of preferences and of decision. The problem, in the form in which it is usually posed, involves two acts – one being the first attempt at getting the grapes, the second being the act of walking away – between which the fox has changed his mind. Note that, by posing the question this way, we are treating its *dynamic* version. There is also a *counterfactual* version of the phenomenon, where the comparison is not so much between acts in two successive situations, but between an actual act and a counterfactual act, which the agent would choose in a counterfactual situation. Sour grapes is involved in the fact that the fox chooses to walk away in the actual situation, where the grapes are difficult to obtain, *although he would have attempted to grab them had they been easy to obtain* (ie. in some counterfactual situation where they were easier to obtain). Although the bulk of this paper will concern the dynamic version, the main points will apply equally to the counterfactual version; where it is not clear how, this will normally be indicated. The advantage of posing the problem in a dynamic form is that it involves a real, observable phenomenon, viz. the fox's decision not to pursue his attempt to obtain the grapes. Even the most behaviourist

perspective on decision making accepts this part of the phenomenon. From this point on, however, things become rather murky. Here are a number of questions or aspects about the phenomenon which are particularly relevant.

What has changed? Under the assumption that the agent is minimally rational, in the sense that he chooses the actions he prefers most, the *preferences* of different available *actions* (attempting to grasp the grapes, walking away) must have changed. In the vocabulary of basic decision theory, in which a large portion of our discussion shall be couched, one would say that his *expected utility* – the utility or "profit" or "pleasure" he would expect to obtain from an action given his beliefs about the state of the world – has changed.[1]

However, this change is not in itself the interesting factor in the sour grapes phenomenon: expected utility changes are widespread, and often easily dealt with, for the simple reason that the expected utility can change for two reasons. On the one hand, such a change can result from a change in the agent's *beliefs* about the state of the world; on the other hand, it could also be a consequence of a change in the amount of "pleasure" or "profit" the agent would obtain if certain consequences were realised, or, to put it in decision-theoretic terms, a change in the agent's *utility*.[2] It is this latter sort of change that sour grapes is often taken to involve. Indeed, this interpretation of the phenomenon is necessary for the role to which it is often put in Social Theory. Utilitarianism relies on the preferences of the different members of the society; sour grapes or "adaptive preferences" is taken to show that these preferences may not be sufficiently stable, thus weakening the utilitarianist position, at least as it is traditionally stated.

The difference between utility and expected utility thus turns out to be very important to the understanding of the sour grapes phenomenon.[3] The difference is usually summed up in the following way: utility is a function on or property of the consequences of the agent's actions, whereas the expected utility is a function on or property of his actions. However, the important difference for current purposes is rather the following: utility is *pure* in so far as the calculation of the utility of a consequence does not depend on the beliefs of the agent, whereas expected utility is *mixed* in so far as the calculation of the expected util-

[1] The discussion of this paper shall mainly be couched in the terms of classical decision theory, where the rational agent is taken to maximise expected utilities. More advanced, non expected utility theories, shall not be considered.

[2] Throughout this paper, there will only question of changes in the *utility functions*, which will be referred to using the expression change of "utility", or at times, in order to help readability, "utilities". Changes in utility values can be derived from the changes in the utility functions.

[3] Often, in discussions of preference change, especially among logicians, this distinction is not drawn (van Benthem and Liu, 2007; Hansson, 1995). In this paper, the theories proposed by such authors shall be interpreted as theories of utility change rather than expected utility change; this interpretation, although debatable, at least allows us to consider to what extent these theories can tackle the problem of sour grapes. Furthermore, such discussions are often couched in terms of preference "update" or "upgrade". In the literature on belief change, update is generally considered to involve changes in belief which accommodate new information regarding changes in the world, whereas "revision" is the name given to changes in belief to accommodate new information highlighting previous errors (Katsuno and Mendelzon, 1992). Adaptive preferences involve changes in the preferences of agents, not changes in the world. For this reason, the discussion will generally be couched in terms of preference or utility *revision* or *change* rather than update.

ity of a particular action depends, as well as on the utilities of the consequences, on the agent's beliefs about the state of the world. If one draws the line so, then approaches to so-called "utility change" which model "utility" as depending on beliefs and which account for utility change in terms of changes of beliefs should be considered as models of *expected* utility change (or, at least, mixed utility change), and not of (pure) utility change.[4] If the fox turns away, as in La Fontaine's version of the fable, saying to himself that the grapes are green, this can be seen as a change of belief: only his *mixed* preferences have changed – the desirability of the grapes depending not only on the general desirability of grape–properties, but also on his beliefs as to whether these grapes have those properties – whereas his *pure* preferences have remained constant – he still enjoys non-green grapes as before. As noted above, a strong intuition dictates that sour grapes involves pure as well as mixed utility change (Elster, 1983, p123). *If* this intuition is correct, then these models provide inadequate accounts of the phenomenon.

Why has it changed? Even if one accepts that sour grapes involves (pure) utility change, it has to be accepted that this utility change is a *consequence* of some change in the situation. As Elster (1983, pp121-2) points out, one might draw a distinction, as to the source of such changes, between those caused by changes in the world – the "state-dependence" of preferences – and those caused by changes in the options open – "possibility dependence".[5] Almost immediately he qualifies the distinction by noting that, given the possible interdependence of *states* and *options*, and the difficulty in getting a clear separation of the two notions, it may be practically impossible to apply this distinction correctly in practice. One might expect that a proper account of sour grapes take account of this distinction and its instability (or, if you prefer, flexibility).

An important point about dependence relations between utilities, states and possibilities, perhaps not clearly kept in focus in Elster's discussion, relates to the *diachronic* or *synchronic* nature of the dependence. Decision theorists such as Karni and Drèze have proposed theories of *state-dependent utility* – where utility is not only a function of the consequence, but equally of the state which the act will "take" to that consequence. This framework is designed to permit a notion of belief differing, roughly speaking, from the behaviourist notion favoured by Savage (Karni, 1996). It describes the decision situation *at a given instant* in such a way that the utility depends on the states: this is *synchronic dependence*. On the other hand, even if the decision situation were such that the utility showed no dependence on states (at that instant), there could be *diachronic* dependence, in so far as the *previous* states, choices and so on have an influence on the *current* utility. This seems closer to the sort of dependence that Elster had in mind, especially given that the notion of possibility-dependence can only really be understood diachronically,[6] as the fact of moving into a situation where only those

[4] Cyert and DeGroot (1975); de Jongh and Liu (2006) are examples of such models.

[5] Given that the definition of a decision problem incorporates the set of options on offer, the second sort of change is essentially a change in the decision problem.

[6] Recall from footnote 5 that a possibility-dependent change corresponds to a change in the decision

possibilities are on offer.

What is the nature of the change? Elster (1983, Ch III) emphasises that the change in preferences is of a causal nature, and may not be intentional on the part of the agent. Whatever the fox's opinions on the change, it was *caused* by his experience of the first attempt at obtaining the grapes; he *did not decide* to change his preferences in the face of this experience.

How fast and how permanent is the change? Elster (1983, pp112-114) takes pains to emphasise that the adaption of preferences in situations involving sour grapes is in principle *reversible*, after a further change in the situation. In fact, intuitions regarding the question of reversibility or permanence of the change differ depending on the time-scale involved. In general, there are three basic intuitions. The first dictates that the fox does not instantly and immediately change his mind about the grapes, and so would take them if the possibility arose immediately after his exclamation that they were no longer desirable. According to such an intuition, "reversibility" of the purported change is very plausible at moments close to the situation in question. The second intuition arises from the idea that it does seem possible, over a longer period of time, and perhaps through force of habit, for the fox to *actually acquire the sort of preferences (regarding the grapes) he claims to have*, so that he would *not* take the grapes if offered. Indeed, many pertinent examples of sour grapes generally involve an extended time span over which the preferences of individuals seem to change.[7] A final intuition, that which is expressed by Elster, dictates that even this long-term change can be reversed by a change in situation: given the correct situation, the fox would once again act in accordance with a preference for grapes.[8] A full analysis of sour grapes should be able to account for these factors. Indeed, the reversibility aspects seem to suggest that whatever modified situation the fox finds himself in could somehow "lead back to" a situation similar to the initial one. A natural, if not inevitable, conclusion to draw is that the information regarding the agent's previous state is *somehow recoverable* from his modified state.

While the counterfactual versions of the previous points are easily derivable from the discussion above, the counterfactual equivalent of the point about reversibility and permanence is worth stating explicitly. For the case of immediate reversibility (the first intuition stated above), it is the following: the fox's utility after a failed attempt to get the grapes is not what it *would* be if the attempt succeeded. Like the dynamic version, it involves the idea that the information about other possible situations is recoverable from one particular possible outcome of his first attempt at getting the grapes: the agent's attitudes in the actual situation (where he decides not to try again to obtain the grapes) are constituted in such a way that information regarding certain counterfactual situations (where

problem, and thus is hard to model synchronically, at least without abandoning the traditional decision-theoretic framework.

[7] See, for example, the case of change of preferences for city or countryside life in Elster (1983, p112 *sq.*).

[8] In Elster's example, someone who moves to the city may acquire a preference for city life, which may be reversed if he moves back to the countryside for a considerable period.

the grapes are easier to obtain) is, at least partially, recoverable. There is no natural counterfactual analogy for the long-term "endurance" of the purported utility change (second intuition); the counterfactual version of the long-term reversibility (third intuition) is basically the same as that stated above.

These considerations indicate the importance of the phenomenon of sour grapes for decision theory. Sour grapes poses the problem of the identity, stability and variability of the central notions of decision theory – utility, belief, expected utility. The basic question of sour grapes is: what changes? Beyond the preferences on actions (expected utilities), is there a change in the preferences on consequences ("pure" utilities), or just a change in beliefs?

In Section 2, three analyses – or, more accurately, analyses schema – for the phenomenon of sour grapes shall be proposed. For each analysis, the extent to which it accounts for the points made above shall be discussed; this counts as a preliminary test of the analyses, which may be used to compare them. However, in the Section 3, it will emerge that the general question posed by sour grapes can be approached in two ways, and that, although the analyses are all consistent with both conceptions of the question, the suitability of the analyses will depend on the precise task which they are supposed to fulfil. One of the approaches to the problem shall be defended against the other, and the analyses shall be judged according to their ability to fulfil the tasks demanded by this approach. Before embarking on the presentation of the models of sour grapes, a few preliminary methodological remarks are in order.

Methodological preliminaries As has been suggested by the discussion so far, the basic framework of Savage-style decision theory shall be assumed here (Savage, 1954). A decision problem consists of a set of states of the world, a set of possible consequences, and a set of acts, considered to be functions taking each state of the world to a consequence. The agent's beliefs are represented by a probability function on the states of the world, his utility (or desirability) of consequences are represented by a utility function (assigning to each consequence a real number, considered to be its utility value). His preferences over acts are thus represented by his expected utility – the sum of the utilities of the consequences of the acts weighted by the probabilities of the states which would yield that consequence – in the sense that he agent prefers an act to another if and only if the expected utility of the first exceeds that of the second.[9] This choice of framework is not intended to imply any particular commitment to this fashion of theorising; rather, its purpose is to clarify the discussion.

Two properties of the traditional Savage framework should are to be noted. Firstly, it applies synchronically: to a particular decision problem, at a particular moment. Secondly, it involves *state-independent* utilities, that is, utilities which are functions solely of the consequences of actions. Since Savage's original work, many have found reason to weaken this supposition on utilities. Indeed, authors such as Drèze and Karni have

[9]One of the issues that will emerge in the final section of the paper is how one comes to fix the beliefs and utilities of the fox before and after his attempt at obtaining the grapes. For the purposes of the presentation of the analyses in Section 2, it may be assumed that the beliefs and utilities in question are unique (and, if one likes, that they have been elicited uniquely, with the aid of whatever representation theorem one prefers); extended discussion will be reserved for Section 3.

developed decision theories which resemble that of Savage, except that the utilities are *state-dependent*: they are functions both of the consequence of the action and of the state which the action "takes" to that consequence.[10] As noted above, these decision theories remain synchronic, in the sense that they apply to a particular decision problem at a given moment. However, they may be seen as more general than Savage's original framework, not only in so far as the state-independence of utility is considered by some to be doubtful, but in the sense that state-independent utilities can be seen as a special case of state-dependent utilities (where the dependence is effectively null). For this reason, the analyses proposed in the following section will be presented in their *state-dependent* version: the utility functions will take as arguments states as well as consequences.[11]

Finally, let us emphasise the importance of the distinction between the point of view of the modelee – the agent – from that of the modeler – the decision theorist. For example, the fact that the modeler elicits certain probability and utility functions representing the beliefs and utilities of the agent does not imply that the agent himself will recognise these as his beliefs or utilities. Indeed, the division of aspects into those of which the modelee is conscious and those which are transparent only to the modeler can be made with respect to different aspects of the model, and indeed yield different interpretations of the same model.[12] So, for example, a common intuition regarding sour grapes, which shall be exploited below, states that the agent does not *really* change his utilities, he only represents the situation to himself in this way. Such an interpretation obviously relies on the distinction between the agent's representation of his mental states and the modeler's.

2 Three analyses

This section contains three analyses – or, to the extent that they require leave precisely specified blanks to be filled by particular mechanisms, three *analysis schema* – for sour grapes. Throughout the section, S shall be the set of pertinent states of the world, where in each state factors such as the position of the grapes and the height of the tree, are determined. C shall be the set of consequences, and shall contain two elements – obtaining the grapes or not obtaining the grapes. The sour grape story involves two situations: the situation before the fox's first attempt at getting the grapes (which shall be called "the first situation"); the situation after this attempt, and in which he takes his decision to try again or to give in (the "second situation").

[10] See, for example, Karni et al. (1983); Karni and Mongin (2000); Drèze (1987).

[11] A disadvantage of this generality is the difficulty in comparing utilities involved in decision problems which do not share the same set of states of the world. This shall become an important consideration in Section 3; for the rest of the discussion, this complication shall be put to one side (situations will be assumed to have states of world which are sufficiently similar to allow comparisons of utilities).

[12] There may be philosophical difficulties with this distinction. However, it is necessary to make it where possible for *methodological* reasons: to avoid confusions regarding the interpretation of the model. In practice, it shall not be difficult to draw the distinction where necessary for the purposes of the following discussion.

2.1 Pure utility change

The simplest analysis of sour grapes takes it at face value: the difference between the two situations is indeed a change in (pure) utilities. In such a model the modeler and the agent (the fox) agree that the fox's utility function is different in the second situation, with respect to the first, and this explains his decision not to pursue his attempt to obtain the grapes.

Writing this formally, let the initial preferences of the fox be determined by probability p_1 and utility u_1, so that the expected utility of an action f is $\sum_{s\in S} p_1(s)u_1(s, f(s))$. Then, according to this model, the probability in the second situation will also be p_1, but the utility will now be u_2. The preferences of the fox (over actions) will thus be represented by $\sum_{s\in S} p_1(s)u_2(s, f(s))$. A full model results when one adds an account of the change from u_1 to u_2; one might expect theories of preference change for example to provide such accounts.

Let us consider how this form of analysis fares with respect to the points raised in Section 1. Concerning the question of what has changed, it is the (pure) utility which is taken to change in this situation. Indeed, the interpretation of this situation as a utility change is unavoidable, unless certain special conditions are satisfied. That is to say, one *could not* even rewrite the representation in the second situation $\sum_{s\in S} p_1(s)u_2(s, f(s))$ *as if* it consisted in a change of belief with a fixed utility (ie. in the form $\sum_{s\in S} p'(s)u_1(s, f(s))$), unless there was a particular relationship between the two utilities: namely, $u_2(s, c) = p''(s).u_2(s, c)$, where p'' is a function on states satisfying the ordinary axioms for probability. Given that this is a rather special case, one can justifiably say that this is essentially a model under which sour grapes really involves a change in utility.

Consider the question of the source of the change (state- or possibility-dependent). Under this analysis, the same possibilities (actions) are available in the first and the second situation, so the change cannot be thought of as possibility-dependent. Indeed, most of the current theories of preference change keep the same options (ie. possibilities) but alter the preferences on them;[13] to this extent, they could only be understood in terms of state-dependent change. As such, there is a general failure to account for the tight relationship between state- and possibility-dependent change. Furthermore, for current theories of preference change, even the interpretation as state-dependent change is not immediate. These theories usually consider changes in the face of statements specifying particular preference relations to be accepted,[14] and it is not trivial to translate the changes in the world – the fox's experience of his first attempt at getting the grapes, say – in terms of such statements. Indeed, these theories seem at best to be modelling *intentional* preference change, as witnessed by the motivation they draw from models of apparently intentional processes (learning from observation, accepting an announcement). This is at odds with the nature of the change involved in sour grapes and noted in Section 1: the fox's experience *causes* the change, he does not *decide* to

[13] Cyert and DeGroot (1975); van Benthem and Liu (2007) and the revision and contraction operations in Hansson (1995) are some examples.

[14] In Hansson (1995), the agent "learns" that a certain outcome has a certain desirability, and alters his preferences accordingly; in van Benthem and Liu (2007), an agent is told to prefer a certain outcome, and alters his preferences accordingly.

change his utility after the experience. If a theory of preference change is to be had which completes the basic framework in such a way as to account for the aspects noted in Section 1, more work is required.

Furthermore, such a completed model would equally have to deal with the reversibility and permanence properties of sour grapes, an aspect on which the analysis, at least *primae facie*, does not fair well. On the positive side, this analysis does account the instance where the subject actually acquires the utility in the long term, modelling it as a straight utility change. However, the modalities of the change seem to have been reversed: an adaption of the utility over a long period is captured here by a sudden revision at a particular moment. Furthermore, it is not certain that the analysis can account for the short- or long-term reversibility of the change, because there is no guarantee that the utility change it proposes is reversible. Indeed, most models of preference change, based as they are on models of belief revision or public announcement, do not allow one to recover the original preferences: like the models they are based on, the previous preferences, and thus other possible preference revisions, are lost when a revision is performed. The reversibility and permanence phenomena thus pose a supplementary challenge for anyone seeking to take up and defend this analysis, which does not seem to have been satisfactorily met by current models of preference change.

This first analysis of sour grapes is firmly embedded in an as yet not fully developed theory of utility change. At this stage, all that can be noted is the inadequacies of the current models of utility change, and the difficulties which an eventual model should overcome. Yet, even under this meagre construal, some may already find the analysis inadequate. All that theorisation of utility change would bring is a theorisation of the particular utility changes occurring in sour grapes. However, it would not explain the particularities of the phenomena, such as the intuition that the fox's utilities do not *really* change, at least not immediately after his failed attempt at getting the grapes. The second model takes this as its guiding intuition.

2.2 Self-justification

An important intuition about the sour grapes phenomenon, briefly mentioned in Section 1, is that it does not effect so much the *action* of the agent (at the moment of the sour grapes phenomenon), but the *way he justifies* or *rationalises* the action (to or for himself). The fox walks away from the grapes in any case; it is the reason he gives himself for walking away that is at issue. Under this interpretation of the phenomenon, although it does not (directly) effect concurrent behaviour, the rationalisation he constructs for himself will effect the utilities and the beliefs *he sees himself as having*.

One would thus want an *internal* model, representing the utilities and the beliefs the fox sees himself as having; this will differ from the representation that an external observer would make of the situation. For someone inclined to think of sour grapes as simple self-justification, the change in the expected utility is properly thought of as a revision of beliefs with information learnt during the first attempt at getting the grapes: he learns that the grapes are more difficult to obtain than previously thought. However, according to such an analysis of sour grapes, the fox represents the change to himself as a change in his preferences for the grapes; that is, as a change of utilities.

As in the previous example, let the initial preferences of the fox be determined

by probability p_1 and utility u_1, so that the expected utility of an action f is $\sum_{s\in S} p_1(s)u_1(s, f(s))$. Furthermore, assume that $p_1(s) \neq 0$ for all s: this assumption captures the fact that the fox does not have any preconceptions about the position of the grapes and the like. It is supposed that the modeler and the fox agree on the fox's probability and utility functions: that is, they represent the fox's state in the first situation, *both for the fox and for the modeler*. The modeler and the fox will disagree however on the representation of the fox's state after his attempt at getting the grapes. For the modeler, the effect of the first attempt can be represented as a change of probability to a new function p_2: thinking of it this way, the fox learns from his first attempt (that it is more difficult than he thought to get the grapes). The change from p_1 to p_2 can be modelled by whatever belief change mechanism one prefers, and different mechanisms will lead to different "completions" of this analysis. After the change, the fox's expected utility thus becomes $\sum_{s\in S} p_2(s)u_1(s, f(s))$. However, as opposed to the case of the previous analysis, it *is* always possible to rewrite the expected utility formula *as if* there was a change in the utility and *not* in the probability. One obtains the representation by $\sum_{s\in S} p_1(s)u_2(s, f(s))$, where $u_2(s, c) = \frac{p_2(s)}{p_1(s)}u_1(s, c)$.[15] This is the sort of change that the fox *sees himself* as making: he has not learnt that the grapes are harder to obtain, he has just changed his mind about whether he wants them or not.

Let us consider how this analysis fares with respect to the points raised in Section 1. Concerning the question of what has changed, the point of view taken on the situation is crucial. All are agreed that the expected utility (preferences on acts) has altered; however, whereas the theorist's representation traces the change to a change in beliefs, the fox represents the change to himself as stemming from an alteration in his utilities. Under this analysis, sour grapes does not pose a *specific* problem for the modeler: it can be modelled with ordinary belief change apparatus. Sour grapes is merely a phenomenon of self-justification, and, at this stage at least, only a change in the representation of one's own preferences.

Turning to the properties and the source of the change, there are two aspects (from the modeler's point of view): firstly, the experience of the first attempt at obtaining the grapes causing a change in beliefs, and secondly, a reluctance to recognise the change in expected utility as ensuing from a change in beliefs. Given that neither of these factors are intentional in themselves (the first attempt is intentional, its result, and the belief change caused, is not), the change comes out as causal rather than intentional. However, as in the analysis presented above, some belief change mechanisms may not be applicable in this case if they cannot be interpreted as modelling "unintentional" belief change. The role of the first attempt – which, in the second situation, can be considered as a state of the world (the world has the property that the first attempt was a failure) – indicates that there is *diachronic* state-dependence, rather than possibility-dependence. Indeed, the fact that the same states and consequences are involved, and thus the same actions are on offer, in the first and second situations, implies that this analysis cannot account for the cases where sour grapes is caused by possibility-dependence, nor for the subtle relationship between state- and possibility-dependence. The diachronic dependence is understood in so far as the belief change mechanism employed yields an understanding of the effect of the fox's negative experience at obtaining the grapes

[15] This is well-defined since $p_1(s) \neq 0$ for all s.

on his subsequent beliefs. Furthermore, the diachronic state-dependence has a synchronic counterpart, in the fox's rationalisation: the sort of state-dependence of the utility in the first situation (ie. in the utility u_1) necessarily differs from the state-dependence in the second situation (ie. in u_2). This difference is particularly clear in the case where the former utility is state-independent: the latter utility will nevertheless be state-dependent, which attests to a form of synchronic state-dependence in the wake of the first attempt at getting the grapes. From the fox's point of view, the synchronic state-dependence in the second situation is an admission that, had the grapes been more easy to get to, they would be more desirable. Depending on whether one considers that the fox has changed his preferences regarding the grapes in general, or only his preferences regarding grapes which require a certain effort to obtain, one might count this as more or less of a weakness of this analysis.

Finally, regarding the reversibility and permanence properties of the sour grapes phenomenon, this analysis does not necessarily fair too well. The probability changes are not in general reversible; indeed, most belief change mechanisms are irreversible. It may thus not be possible to recover the original utility from the new utility the fox thinks of himself as having. However, the fact that the utility has not changed, under the modeler's account of the scenario, may go a fair way to explaining why the fox would accept the grapes if offered immediately. In this sense, the model may be able to account for short-term reversibility. On the other hand, the entire dependence on a change in belief does not seem to explain the apparent endurance of a utility change over long time periods, which was noted in Section 1. Similarly, the phenomenon of long-term reversibility (reversibility after an enduring utility change) seems difficult to account for only in terms of belief change.

This analysis perhaps fairs better than the previous one – for one thing, there is a larger amount of work on belief revision to draw upon – but it still has trouble accounting for some of the subtler properties noted in Section 1. Moreover, there are several other aspects of the model which may count as unsettling. There is a certain intuition according to which *there is no belief change involved in sour grapes phenomena*. The fox knew the position of the grapes, the height of the bush and similar information *before* his first attempt: so what has he learnt? The natural answer seems to be that he learnt the chances of success at obtaining the grapes, given that they are at such a height. As is infamous in the domain, distinctions of beliefs from other aspects depend on the way decision problems are set up: in the current case, the states of the world pertain entirely to the situation of the grapes (and, perhaps, physical facts about the fox), and no information on these issues is apparently learnt in his first attempt at obtaining the grapes. The chances of him succeeding correspond rather to the chances that the acts on offer (which are, technically, functions from states to consequences) *reliably* effectuate the transitions from state to consequence which they claim to. This is the guiding intuition for the third analysis.

2.3 Reliability of acts

Under the final approach to the phenomenon of sour grapes, the fox does not learn anything about the states of nature, nor does he alter his utility on consequences, but *he does change his opinion about the reliability – or if you prefer, the chances of success*

– *of the acts on offer*. This change can be absorbed into the utility function and thus be interpreted as a change in it; on the other hand, it may be left explicit, where it receives a natural interpretation.

As in the previous analyses, let p_1 and u_1 be the probability and utility of the fox before his first attempt at obtaining the grapes, so that his preferences in the first situation are represented by $\sum_{s\in S} p_1(s).u_1(s, f(s))$. As a result of his attempt at obtaining the grapes, the probability and utility functions do not change, but he places a (non-trivial) *reliability measure*, γ, on the pairs (s, c) ($s \in S, c \in C$). For an act purporting to send s to c, $\gamma(s, c)$ measures the *chances of success* of the act, when the state of the world is s.[16] Despite the vocabulary used, there seems little reason why γ should be a probability measure, although, assuming the ordinary boundedness assumptions, it can be normalised so as to take values between 0 and 1 (0 meaning that the act will certainly fail, 1 that it will certainly succeed). The representation of the posterior preference relation (in the second situation) will thus be given by $\sum_{s\in S} p_1(s).\gamma(s, f(s)).u_1(s, f(s))$.

Regarding the questions posed in Section 1, note firstly that, like the first analysis (Section 2.1), it is not always possible to reformulate the change in his preferences as if it consisted in a change in belief.[17] However, it *is* always possible to reformulate it as if it were a utility change: $\gamma(s, c).u_1(s, c)$ may be thought of as a utility function. In this sense, the experience can be thought of as causing a *change in the utility function* to $\gamma(s, c).u_1(s, c)$; what is more, γ characterises exactly this change. Evidently, it is *not* necessary to see this as a utility change, because $\gamma(s, c)).u_1(s, c)$ is not the only utility involved in second situation: there is still the initial utility u_1. u_1 is, so to speak, the *pure* utility, independent of the situation,[18] whereas $\gamma(s, c).u_1(s, c)$ is the utility *in this situation* – relative to the situation in so far as the situation limits, through γ, the *accessibility* of c, or the chances of actually obtaining this consequence. Many of the attractive properties of the analysis come from the availability of these two utilities. The first of these properties has already been evoked: in this analysis, sour grapes comes out as a *change in the situation-indexed utility*, though not in the *pure utility*.

Indeed, the interpretations of these two utilities depend on the view one has of the agent. If the agent is considered to be conscious of the aspects of the model, then he is perfectly lucid about the influences on his preferences over acts which arise from his utilities, and those influences which arise from the chances of success of the different acts offered. When the fox mumbles that the grapes are no longer desired, his affirmation is only true – and only meant – in the context of the current situation, since it refers to the situation-dependent utility rather than any more stable or transcendental utility. However, if the model is seen as only that of the modeler, the fox may not be so clear about the difference between the two utilities. This confusion has consequences for the appropriateness of the model, and notably the distinction between the two utilities. He might consider a utility change to have occurred, whereas the "underlying" utility (the

[16] For the purposes of questions regarding uniqueness and representation theorems, γ may be taken to be a technical translation function which engineers would reveal by experiment (for example, by offering appropriate bets on the success of acts).

[17] This will only be possible in the degenerate case where γ is independent of its second value c, and where it is a probability measure on S.

[18] Recall footnote 11: complexities regarding situations which do not share the same states of the world are left aside here.

u_1 retained by the modeler) has not actually changed.

This duality of utilities has consequences for the question of whether the change is to be considered as state- or possibility-dependent: it allows the analysis to be faithful to the flexibility of this distinction, noted in Section 1, and left unaccounted for by the previous two analyses. On the one hand, the change between the two situations (before and after his attempt at obtaining the grapes) can be considered as a change of *state* (to a state where he has had the experience of his attempt) as in the previous two cases; there is diachronic state dependence. On the other hand, the function γ represents limits on the range of actions he can expect to carry out *successfully*; in other word, it limits on the *possibilities effectively* available to the fox in second situation. In this sense, there is diachronic possibility dependence: the experience of the first attempt causes the introduction of the factor γ, which can be thought of as restricting possibilities. Moreover, this diachronic subtlety is reproduced on the synchronic level. On the one hand, there is possibility dependence if one considers the agent as keeping his utility u_1, but having his range of actions restricted by the reliability function γ. On the other hand, there is state dependence if one takes his situation-indexed utility $\gamma(s,c).u_1(s,c)$, in so far as this utility has a different dependence on states from the utility in the first situation (u_1).

Furthermore, the duality of the two utilities, and their different behaviour across situations, may help to account for the reversibility and permanence phenomena. $\gamma(s,c).u_1(s,c)$ is the utility in the second situation, but u_1, the pure utility, is still conserved as a separate term, which may be extracted, used and discussed. For example, the immediate reversal of preferences may be understood as reverting back to the pure utility u_1; this can equally be considered as a change in γ (corresponding to the shift to another situation). Here, the fact that the pure utility u_1 remains constant between situations, whereas the situation- or context-dependent parameter γ varies, is exploited; indeed, in many cases, the relationship between the second situation and other situations may be understood in this way. On the other hand, the fact that the analysis offers another utility, the agent's *situation-dependent* utility $\gamma(s,c).u_1(s,c)$, which the fox may mistake for his utility proper, may be able to play a role in understanding long-term utility change. If the agent remains for an extended period with a generally similar set of acts, and a similar reliability measure on these acts, this utility will continue to apply (the same factor γ applies to these acts), and eventually, as he comes to presuppose that these are the only acts on offer, and that they have these chances of success, this utility will completely usurp the pure utility u_1. Long-term reversal of preferences could be understood in terms of a situation change which is drastic enough to invalidate this presupposition, by introducing new acts or by changing the reliability measure γ, in such a way that resort to the pure utility is once again required.[19] Under this account, long-term reversibility is a similar sort of effect to short-term reversibility, though of differing degree.

Just as for the other analyses, this is more a sketch than a fully developed analysis. A full analysis would require an account of the appearance and dynamics of the factor γ. Such an account should make clear its dependence on the result of the fox's

[19] Indeed, in Elster's example of long-term reversibility (of the preference for city life; see footnotes 7 and 8), a drastic change is required (a move to the countryside for a considerable period).

first attempt at getting the grapes (the causal aspect of the change[20]), as well as, more generally, the relationship between the functions γ in different situations (related to the reversibility), and an account of the possible absorption of γ into the agent's utility over time (related to the long-term permanence of the change). However, the introduction of γ permits two notions of utility which, to a certain point, reconcile the intuitions that the utility changes, that it has not "really" changed at the moment when the fox walks away, but that, over time, it may really change. Indeed, the duality of utilities allows the analysis to take account, in a way the previous analyses do not, of the subtler aspects of the sour grapes phenomenon, such as the flexible distinction between state- and possibility-dependence, and the question of reversibility and permanence. The price to pay, if it even seen as such, is the introduction of a factor γ and a representation of preferences which, to the knowledge of the author, has not appeared previously in the literature.

3 Getting your teeth into sour grapes

Three possible analyses – or rather, analysis schemas – of the phenomenon of sour grapes have been presented; for each, their abilities to account for the properties of the phenomenon noted in Section 1 have been discussed. One might attempt to decide between these analyses on the basis of their performance with respect to these properties. A more thorough comparison, however, would only be possible if it were made clear what exactly is expected of a model or analysis of sour grapes. In this concluding section, it is the general question of what one expects in general from a model of dynamic phenomenon, such as sour grapes, which will be at issue. Two strategies for dealing with such phenomena will be considered, which conceptualise the problem in different ways, and demand different things from analyses of the phenomenon.

As noted in the introduction, the basic question of sour grapes is determining what changes: beliefs or utilities.[21] This involves two sorts of problems. One is the problem of distinguishing the agent's beliefs from his utilities *in a given situation or decision problem*. The other is the problem of studying the relationship between his beliefs and utilities *in different situations* (be they subsequent or counterfactual). Much work has been dedicated to the first problem taken in isolation from the second, and indeed on the second – at least for the case of belief change – taken in isolation from the first. However, given that both problems are central to sour grapes, the challenge is to *account for both of them*. A model of sour grapes should permit a distinction between the agent's beliefs and his utilities in a given situation *and* deliver an understanding of their behaviour in time or under changes of situations. Only such a model would be able to say something about what changes are actually involved in the phenomenon of sour grapes.

There are two strategies for dealing with this pair of problems. One strategy, the

[20]Under this analysis, the change is of a causal nature, brought about by the fox's experience of his first attempt at obtaining the grapes. As in the previous cases, the fact that it is not necessarily intentional may place restrictions on the form of the theory of γ which would complete this analysis.

[21]There could be a change in both: this would correspond to a certain mix of some of the analyses given above, and not treated separately here.

most common, proceeds by *first* sounding out the beliefs and utilities in each, individual situation, *independently* of other situations, and *then* treating the question of how the beliefs and utilities have changed from one situation to the next. An alternative strategy considers such a factorisation impracticable or undesirable, and so allows the decision about the beliefs and utilities in a particular situation to depend on other (counterfactual or subsequent) situations. It deals with the problem of distinguishing beliefs and utilities in a given situation and that of understanding their changes between situations *simultaneously*, in such a way that the conclusions relative to one make reference to observations regarding the other.

Note that, at least in the dynamic version (as opposed to the counterfactual one), the second strategy is as behaviourist as the first. It still refers solely to the actions of the agent (and his preferences over possible actions), although the actions involved are spread out over a larger time-scale. Furthermore, the analyses proposed in the Section 2 and the points made about them do not depend on the adoption of one strategy or the other. Indeed, they can be understood equally well, albeit slightly differently, under both strategies. If the former strategy is preferred, then one will suppose that, given appropriate conditions, a unique representation in terms of beliefs and utilities can be obtained:[22] it is the changes in these beliefs which the analyses purport to capture. On the other hand, if the latter strategy is assumed, then the analyses present the structure of the changes of beliefs and utilities, *without* the supposition that these beliefs and utilities can be provided separately from the question of the changes they undergo from one situation to another. Although all the analyses are compatible with both strategies, it may nevertheless turn out that some of the analyses are better suited to the demands of a particular strategy than others.

In this final section, we will mainly be concerned with the second strategy: if any phenomenon can argue for the second strategy over the first, sour grapes, as a dynamic phenomenon involving precisely the question of the distinction between beliefs and utilities, can. Several disadvantages of the first strategy shall be raised. These will suggest that the *second* approach is *primae facie legitimate*. The article will close with several remarks about how it could be applied, and a discussion of the consequences of adopting this strategy for the adequacy of the preceding analyses.

3.1 The common strategy

The first proposed strategy consists in eliciting the beliefs and utilities of the fox before his first attempt at obtaining the grapes and his beliefs and utilities after this attempt *separately*, by making reference only to the respective situations. Under this perspective, the analyses proposed above can be empirically tested, in so far as it can be checked whether the behaviour of the beliefs and utilities are as described by a particular analysis. For such a strategy to be viable, there needs to be a way of sounding a *unique* set of beliefs and utilities for the agent (a unique pair of probability and utility functions) in each situation: without sufficient uniqueness, no non trivial answer can be given to the question of what has changed. To do this, one generally employs the so-

[22] As discussed below, this will normally be done with representation theorems. Given that it is not the purpose of this paper is to discuss the technical details, we omit the relevant axiomatisations and theorems.

called *representation theorems* of decision theory. Potential weaknesses in this strategy arise, firstly, from the conditions under which beliefs and utilities can be (uniquely) specified in a given situation using representation theorems, and secondly, from the consequences of the strategy of eliciting beliefs and utilities whilst totally ignoring their dynamics.

Representation theorems generally work by posing a set of conditions on the agent's preference relation over acts such that, if these are satisfied, the preference relation is represented by a probability function and a utility function which satisfy certain properties, and are *unique* in an appropriate sense. The most famous example is doubtless that of Savage (1954), which, for a preference relation satisfying certain properties, delivers a unique probability function and an effectively unique state-independent utility function[23] representing the agent's preferences. Karni, Drèze and others have proposed conditions which guarantee more or less unique probability and state-dependent utility functions.[24] However, as has been pointed out often, many of the conditions demanded on the preference relation do not seem to apply in actual decision situations. It is not the place to enter into these debates, but only to note that certain of the conditions are particularly unsuitable when emphasis is being placed on decision problems posed in *particular* situations; such is the case when the change in attitudes between situations is at issue. The "structural axioms" of decision theory,[25] such as Savage's P6, demand that the decision situation contain an infinite number of states and that the agent's preferences be fine enough to distinguish between acts which differ only by their effects on a small fraction of these states.[26] For real situations, this obviously does not apply; however, normal responses to this observation, which often involve a recourse to some *global* situation,[27] are not appropriate in the current discussion, since they do not necessarily preserve the dynamic aspects which are at issue here.

The other general objection to the strategy of determining the probability and utility functions in the various situations before studying the (dynamical or counterfactual) relationships between them raises the worry that these relationships may become incomprehensible under this approach. As Karni notes in his critical discussion of Savage's method for eliciting beliefs (1996, p256-7), the beliefs elicited using representation theorems such as Savage's in a particular situation is at odds with the agent's behaviour in related situations.[28] Supposing that Savage's theorem is applicable in these other situations, this objection would seem to suggest that there is *no comprehensible relationship between the beliefs and utilities elicited by means of the theorem in the two situations*. This conflicts with the intuition that the two situations are strongly

[23] Namely, it is unique up to affine increasing transformation. Such subtleties may be ignored for current purposes.

[24] See the references in footnote 10.

[25] As Suppes (1984) calls them.

[26] Theorists who deviate from Savage's framework generally require equally strong assumptions: for example, the framework proposed by Anscombe and Aumann (1963), and adopted by Karni and Mongin (2000), consists a finite sets of states, but an infinite (and structured) set of consequences.

[27] Savage (1954, p83): "a person has only one decision to make in his whole lifetime. He must, namely, decide how to live, and this he might in principle do once and for all."

[28] Karni considers Aumann's example of the husband betting on his wife's operation (see Drèze (1987, p77)), and in particular the continuation where he advises a friend on a similar bet: the point is that the husband's advice conflicts with the bet he has made.

connected. With such representation theorems, one gains a unique representation of the beliefs and utilities, at the price of even the most basic understanding of how the beliefs and utilities in different situations are related. Karni's objection seems to point to a general criterion for representation theorems: they should not *only* yield synchronic or instantaneous representations of preferences in terms of beliefs and utilities (ie. in terms of a probability and a utility function), but *they should furnish representations which aid in the understanding of cross-situational (dynamic or counterfactual) phenomena involving beliefs and utilities*. This requirement on representation theorems is well-motivated: there is little point distinguishing the beliefs and utilities of an agent involved in a single decision in a particular situation, if one is not also interested in the relationship with the beliefs and utilities involved in other decisions in other situations. For want of a better name, let us call this the *dynamic requirement*.

Although a range of different representation theorems have been proposed (each proposing a different pair of probability and utility functions), it is still unclear which of them deliver the probability and utility functions which are the most fruitful for understanding the behaviour in related situations. Indeed, such theorems are not designed with the dynamic requirement in mind, and, in discussions of the theorems, little attention is paid to whether the representations *can* account for such relationships. This is a second reason for avoiding the use of traditional representation theorems, and more generally the strategy which employs them, in the approach to sour grapes.

3.2 An alternative strategy

The second approach to phenomena such as sour grapes deals with the question of distinguishing beliefs and utilities in particular situations and the question of their changes simultaneously. In judging what the beliefs and utilities of the agent are in a particular situation at a particular moment, it is permitted, if not necessary, to look at his actions, and perhaps beliefs and utilities, in other situations at other moments (or in other counterfactual situations). Such an approach is certainly not without intuitive support, at least for the phenomenon of sour grapes under discussion here. Indeed, this approach is faithful to the adage that, to see whether the fox has *really* changed his mind about the desirability of the grapes, one needs to see how he acts in other situations where they are more accessible (if they were offered on a plate for example, or if, just after turning away, he spots a ladder). The approach to the problem discussed above ignores this intuition, in so far as it claims to elicit the fox's beliefs and utilities *solely on the basis of his preferences in the situation where he chooses to walk away*. Indeed, as noted above, that strategy will generally *fail to account* for intuitions, such as this one, which involve other situations.

By posing the question of the relationship between situations at the same time as that of the beliefs and utilities involved in a given situation, the second strategy takes full heed of the dynamic requirement stated above. Indeed, the task now becomes that of specifying a unique set of beliefs and utilities (probability and utility functions) in each situation in such a way that, not only they represent the preferences of the agent in the appropriate situation, but equally they permit an understanding of the relationship between the agent's preferences in suitably related situations. The sort of understanding required in the last clause remains to be made precise, and probably can only be done

so as different sorts of relations are proposed and debated.

This is a reformulation *both* of the problem of distinguishing beliefs and utilities in a given situation *and* of the problem of sour grapes, which involves the supplementary question of the change in beliefs and utilities. The fruitfulness of this reformulation, and the possibility of successfully tackling these problems by posing them in these terms, involves three factors or suppositions. The first is that the data available cannot be restricted to the situation or situations of direct interest (the situation where the beliefs and utilities are to be elicited; the two situations involved in sour grapes). The second is that there must be some situations where it is relatively easy to elicit beliefs and utilities. The third is that there is a general consistency in the relationship between situations. Let us consider these points one at a time, focussing in particular on the example of the fox's sour grapes.

As noted above, sour grape phenomena generally involve the comparison between *two* distinct situations. Although the state of the agent (and thus of the world) may differ between the two situations, the general aspects of the decision problem which relate to the preferences at issue do not. So, for example, although between the situations before and after the fox's first attempt something has changed, the *states of the world* and *consequences* (respectively, relating to the position of the grapes and whether or not he obtains them) *have not changed* and so *neither have the acts on offer*. Accepting the dynamic requirement, and its demand that the elicitation of belief and utility in a situation should involve appeals to other situations, this pair of situations does not seem to suffice as *data* for the task of identifying the beliefs and utilities involved in each. Indeed, identifying the beliefs and utilities in each situation in a way which depends only on the other situation, to then use these in the consideration of the relationship between the beliefs and utilities in the two situations, seems flatly circular. It is thus natural to bring a third situation into the fray.[29] The data in the sour grape problem is *not* limited to the two situations involved, but to other, appropriately related situations.[30]

The evident question that now arises, and this is the second point to be made about this sort of approach, is that of the nature of this supplementary situation. A situation which shares the general aspects of the decision problem (sets of states of the world, sets of consequences, and thus acts on offer) with the two original situations, though perhaps differing in utility or belief, would be of only marginal help. This is because the same sort of complicated question of belief and utility change between this third situation and the two original ones would arise again, and it is not clear if it would be any easier to resolve. This would seem to argue for a third decision situation which differs more radically from the original two: namely, one which *either does not share their set of states of the world or their set of consequences, and thus, does not offer the same acts*.[31] The interest of this *decoupling* of consequences and states of the world is

[29] In general, more than one supplementary situation will be required, but only one shall be considered for the sake of this discussion.

[30] This is not the first time that people have suggested relying on other situations in the elicitation of beliefs and utilities, although the strategy has not always been formulated in such explicit terms. Karni and Mongin (2000) use "hypothetical objects of choice", which (putting aside complications regarding conditionals) are the acts which the agent would choose in a situation where a certain condition was realised (he chooses to go out, *if* it does not rain, ie. in the situation where it does not rain). More recently Karni (2006) refers to the agent's preferences after Bayesian update (ie. in a subsequent situation), in his elicitation of beliefs.

[31] The existence of such a third situation depends on the fact that the two original situations are not global.

that there follows a decoupling of beliefs and utilities. Consider a third situation which shares, say, the same consequences as the first two, but with acts which operate on a different set of states of the world: the beliefs in the previous two situations are no longer relevant in this third situation. This should aid the comparison of the utilities between the situations, and notably, show whether they really changed. If the fox were offered the grapes on a plate, his reply would indicate whether his utilities *really* had changed, or whether it was only factors relating to the position of the grapes, and the acts required to obtain them, which changed.

It is of course true that similar problems to those mentioned above may affect this third situation: it may not be possible to deduce a unique probability function (on the new states of the world) and a new utility function (on the old consequences). On the other hand, a judicious choice would circumvent this problem. It suffices to choose a situation with a particularly simple set of states of the world (for example), and where appropriate conditions for eliciting utility functions are satisfied, so that the utility and probability functions can be distinguished uniquely. The situation where the fox is offered the grapes on a plate is such an example. Indeed, this sort of method effectively follows in the footsteps of several well-known and useful techniques in decision theory; moreover, although the techniques were not originally presented in this way, they receive particularly plausible interpretations in terms of the implication of another, simple situation. Consider Ramsey's supposition that there is an "ethically neutral statement" (one for which the agent is indifferent as to whether it is true or false; Ramsey (1931)), or Savage's axioms P2, P3 and P4: although these may be controversial assumptions in many cases, it is surely possible to find simple (perhaps counterfactual) situations where they are satisfied, and thus where the results relying on them may be applied.[32]

The final point to be made about this style of approach concerns a final supposition: namely, that between the sour grapes situation and such a third situation, the beliefs or utilities of the agent have not changed. If the comparison with the third situation is to be useful, it is necessary to assume that the utility (or probability) is the same as in the original situation. This supposition is, essentially, the price to pay for adopting this strategy: given that the question of eliciting beliefs and utilities in individual situations is not resolved without recourse to other situations, suppositions regarding the relationships between situations are required. Under the first strategy described above, where beliefs and utilities can be sounded independently of other situations, such suppositions would be empirically verifiable; not so now. In the example taken above, it needs to be supposed that the fox's preferences for the grapes does not change between the situation where he walks away from the grapes and the situation when he is offered them on a plate. Such relations are defendable on a case by case basis: in the example, it seems plausible if the grapes are offered not long after the fox has decided to walk

Without wanting to enter into a long discussion, let us just note that in most informal discourse about decision, and in most applications of the theory, this is not the case: not *all* the aspects that could ever be in play are effectively involved (eg. in the states of the world involved in the fox example, the result of a particular coin toss is left unspecified).

[32]Ramsey's and Savage's assumptions are mentioned here simply as examples; as noted above, it shall not be generally possible to employ their representation theorems as such, because their structural axioms are unacceptable in ordinary situations. Full discussion of the appropriate conditions and techniques for belief and utility elicitation would take us too far from the point of the paper.

away, though perhaps too close or too far from that moment.[33] It is nevertheless worth emphasising why one should expect *in general, for well-chosen pairs of situations*, this sort of supposition to be valid. The supposition basically states that there is a degree of *stability* or *constancy*, a *minimality of change* between situations. However, a supposition of this sort is required for the study of any *relationship* between situations, such as the questions of belief change, of utility change and thus of sour grapes. To talk of change, one must be able to make sense of what it means for there to be lack of change; this is generally the "null" state (of the relationship) from which theories of change begin. It therefore follows that for cases where there is no reason for change, a basic constancy must be assumed; if not, cases of change would end up being chaotic and no theory of change would be possible. The supposition that, in general, there are such stable cases does not count as particularly contentious, especially if the question of change has already been invoked.

3.3 Consequences for analyses of the phenomenon

Adopting the second strategy implies a slightly different set of desiderata for a model and analysis of sour grapes. It cannot *just* suppose the utilities and beliefs given, and concern itself with the changes from one situation to the next; it must also take heed of the fact that, under this approach, the question of change is partially interlinked with the question of the beliefs and utilities in a particular situation. The analysis must not only permit the understanding of the particular change involved in sour grapes (between the situation before and after the fox's first attempt at obtaining the grapes), but *it should do so in such a way that the representations of beliefs and utilities it uses is amenable to the understanding of relationships with other situations*. For example, the belief and utility representation used should be able to partially account for actions, beliefs and utilities of the fox in other, subsequent or counterfactual situations.

The reversibility and permanence phenomena signalled in Section 1 thus turn out to have a particular importance under this approach to sour grapes, in so far as they involve other situations (more or less temporally distant). Note however that the status of these situations, in their relation to the initial situations involved in sour grapes, differ. For a situation involved in short-term reversibility, the supposition that there is more affinity with the utility function involved when the fox walks away is rather plausible, especially given the temporal and thematic proximity. On the other hand, given the temporal distance between the situation where the fox walks away and the situations involved in long-term permanence and reversibility,[34] there is less reason to suppose a common utility function between the them.

The remarks made in Section 2 regarding the performance of the analyses with respect to these phenomena are thus crucial in the evaluation of the appropriateness of the analyses under this approach. Recall that the first analysis (utility change) treats a gradual, long-term change of utility as if it were an immediate occurrence, and a short-term apparent constancy of utility as a serious of violent changes, whereas the

[33] If the example of the offer of grapes on a plate is not to your taste (the fox's pride might prevent him from accepting), consider the case where, just as he walks away, he spots a ladder ...

[34] The more complicated case of long-term reversibility shall be left aside for the purposes of this discussion.

second analysis (belief change accompanied by self-justification), while it gives some hold on the short-term constancy, is not very fruitful when applied to the long-term case. As for the third account (reliability of acts), it offers two utilities, of which one – the *pure* utility – naturally accounts for the relationship with immediately subsequent situations, whereas the other utility – the agent's *situation-dependent* utility – may provide an account of the long-term change in the utility, if the situations continue to offer the same set of acts (and thus support the same function γ). This comparison seems to favour the third analysis. This difference in performance may be interpreted in terms of the way that the analyses deal with situation change, as follows. The first two analyses involve a simple utility change or belief change; there is thus no distinction between those aspects of the phenomenon which change "for good", and those which are specific to the situation where the fox walks away. However, if one accepts that one should compare different situations to determine the beliefs and utilities in each, and that such a comparison will rest on the supposition of stability of beliefs or utilities between certain situations, the identification of *situation-dependent factors* becomes crucial. The third analysis involves just such a situation-dependent factor: namely γ. This is why it promises to be a fruitful framework for thinking about sour grapes, and perhaps any phenomenon where the question of change is at issue, and where this question cannot be easily distinguished from the question of what changes.

Sour grapes poses an intricate knot of challenges, concerning the identification of belief and utilities, and their changes. Though often treated separately, these questions may not prove as detachable as often assumed. This is the main moral of the paper. If one accepts that the questions have to be, to a certain extent, attacked together, then different requirements are placed on purported models. They must not only take account of the change in the agent's attitudes in the sour grapes story, but equally the relationships with his attitudes and actions in other, subsequent or counterfactual, situations. The distinction between the aspects particular to the situation and those which transcend several situations acquires increased importance. The analysis proposed in Section 2.3, based on the idea of a reliability measure on acts, is the second main contribution of this paper. It gives a natural account of the idea that the agent's pure utility, which applies in a range of more or less contemporary situations, differs from the situation-dependent utility, of which he talks when he states that the grapes are undesirable. It allows a more supple understanding of the relationship between the decision problems, the beliefs and the utilities in different situations, of what changes and of what remains the same. That is, it offers not only an account of sour grapes, but equally an account which promises to be amenable to the larger study of the agent's mental states in different situations.

References

Anscombe, F. J. and Aumann, R. J. (1963). A definition of subjective probability. *The Annals of Mathematical Statistics*, 34:199–205.

Cyert, R. M. and DeGroot, M. H. (1975). Adaptive utility. In Day, R. H. and Groves, T., editors, *Adaptive Economic Models*, pages 223–246. Academic Press, New York.

de Jongh, D. and Liu, F. (2006). Optimality, belief and preference. In Artimov, S. and Parikh, R., editors, *Proceedings of the Workshop on Rationality and Knowledge, ESSLLI 2006*.

Drèze, J. H. (1987). *Essays on Economic Decisions under Uncertainty*. Cambridge University Press, Cambridge.

Elster, J. (1983). *Sour Grapes. Studies in the subversion of rationality*. CUP, Cambridge.

Hansson, S. O. (1995). Changes in preference. *Theory and Decision*, 38:1–28.

Karni, E. (1996). Probabilities and beliefs. *Journal of Risk and Uncertainty*, 13:249–262.

Karni, E. (2006). Bayesian decision theory and the representation of beliefs. Working Paper.

Karni, E. and Mongin, P. (2000). On the determination of subjective probability by choices. *Management Science*, 46:233–248.

Karni, E., Schmeidler, D., and Vind, K. (1983). On state dependent preferences and subjective probabilities. *Econometrica*, 51:1021–1032.

Katsuno, H. and Mendelzon, A. O. (1992). On the difference between updating a knowledge base and revising it. In Gärdenfors, P., editor, *Belief Revision*. Cambridge University Press, Cambridge.

Ramsey, F. P. (1931). Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*. Harcourt, Brace and Co., New York.

Savage, L. (1954). *The Foundations of Statistics*. Dover, New York. 2nd edn 1971.

Sen, A. K. and Williams, B., editors (1982). *Utilitarianism and Beyond*. Cambridge University Press, Cambridge.

Suppes, P. (1984). *Probabilistic Metaphysics*. University of Uppsala Press, Uppsala.

van Benthem, J. and Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Classical Logic*, forthcoming.