Chapter 2
Three Analyses of Sour Grapes

Brian Hill

Abstract 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. The question of preference change has been considered by decision theorists and, more recently, logicians. The former phenomenon seems a natural candidate for application of the latter theories. The fundamental question of sour grapes is: what is it that changes – the agent's beliefs or his utilities? The aim of this paper is to consider the replies that decision theorists and logicians can offer to this question. Besides the interest of the phenomenon as a case study for theories of change, it raises two general points. Firstly, besides a belief change and a utility change, there is a third possibility for the source of a given change in preferences: a change in the decision-maker's perception of the choice he is faced with. Secondly, traditional methods for eliciting beliefs and utilities do not function well in cases where several situations are involved and the relations between the agent's attitudes in the different situations are at issue. An elicitation method is sketched which purports to deal more adequately with such cases. Although based on independent motivations, it provides another argument for the importance of taking into account how the decision-maker perceives the choice he is faced with.

2.1 What is Sour Grapes?

2.1.1 The Phenomenon and the Challenge

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 phenomenon, which goes under the name of adaptive preferences or sour grapes, has long caused a stir in Social Theory, mainly because of its importance

B. Hill
GREGHEC, HEC Paris and IHPST. 1 rue de la Libération, 78351 Jouy-en-Josas, France
e-mail: hill@hec.fr.

T. Grüne-Yanoff and S.O. Hansson (eds.), *Preference Change: Approaches from Philosophy, Economics and Psychology*, Theory and Decision Library A 42,

in the debate over utilitarianism (Elster 1983; Sen and Williams 1982). After all, if people's desires change so whimsically, how could any aggregate of these be a proper guide for social choice? But the phenomenon also involves many themes of recent and ancient interest to decision theorists and logicians, such as choice and preference change. The aim of this paper is to see what understanding these theories can offer of the sour grapes phenomenon. More important than the particular challenge posed by this specific phenomenon will be general issues it raises.

Throughout this paper, we will concentrate on the example in La Fontaine's fable. It 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. The challenge of sour grapes is to understand the relationship between the pair of acts and the potentially changing attitudes on which they are based.

An appropriate framework for couching discussion of attitudes, their role in action, and their changes is classical Savagean decision theory. Faced with a choice among several options, a (rational) agent establishes what he considers to be the possible consequences of the options and identifies the facts about the world which affect the realisation of these consequences given that the option is chosen; subsequently, he chooses the option which, given what he believes about the relevant properties of the world, yields on average the most attractive consequences. Formally, the decision problem is represented as a set of possible consequences, a set of states of the world and a set of acts – functions taking each state of the world to a consequence – which are the objects over which the agent is to choose. The agent is taken to have a preference relation over the acts, which determines his choice. We shall call the tuple consisting of a set of states, a set of consequences and a preference relation over the acts taking these states to these consequences a *decision situation*. According to classical decision theory, the attractiveness of the consequences in a decision situation for the agent are represented by a real-valued *utility function*: the more attractive consequences have higher utility. In a similar way, the agent's beliefs about the state of world are represented by a *probability function* on the set of states. These attitudes determine his preferences over acts: the preferences are represented by his *expected utility*, in the sense that he prefers one act to another if the expected utility of the former is larger than that of the latter. Formally: for a probability p and a utility u, the expected utility of an act f is $\sum_{s\in S} p(s).u(f(s))$. The choice of the expected utility framework is not intended to imply any particular commitment to this fashion of theorising; rather, its purpose is to facilitate precise discussion. Indeed, although it shall be assumed that the agent is an expected-utility maximiser, rather than a maximiser with respect to some other more complicated non-expected utility decision rule, many of the points hold for other models of decision (see below for remarks concerning the relationship to Jeffrey's decision-theoretic framework, which is preferred by many philosophers).

To reformulate the sour grapes story in decision-theoretic terms, the fox's preferences for different available acts (attempting to grasp the grapes, walking away) has changed between his first attempt to get the grapes and his walking away. Assuming him to be minimally rational, this means that the expected utilities of these acts must have changed. However, this change is not in itself the purportedly interesting property of sour grapes. Expected utility changes are widespread, and can occur for two reasons: on the one hand, such a change can result from a change in the agent's beliefs about the world, on the other hand, it could also be a consequence of change in his utility. For the use to which sour grapes is put in the debate over utilitarianism (Elster 1983; Sen and Williams 1982), it is crucial that they are not just cases of belief change, but cases of utility change.[1] If there is utility change, so the argument against utilitarianism goes, then the utilities of different members of a society may not be stable; however, utilitarianism relies on these utilities, so if they are unstable, the utilitarianist position, at least as it is traditionally stated, is weakened. For this argument to be valid, it needs to be determined that it is indeed utilities rather than beliefs which change in an apparent case of sour grapes.

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 acts (expected utilities), is there a change in utilities, or just a change in beliefs? From the point of view of a decision-theorist or a logician, this question constitutes the principal challenge posed by the phenomenon of sour grapes.

One can conceive of two methods for answering this question. The *direct* method relies on decision-theoretic machinery: one uses classical techniques to elicit the beliefs and utilities in the appropriate situations and then compares them to see which have changed and how. This method is not open to theorists who do not have a way of eliciting the agent's attitudes on the basis of his preferences; it can be employed by decision-theorists, for example, but not by logicians. An alternative, *indirect* method consists of finding properties of the change involved in the sour grapes story which are *only* possessed by belief changes or utility changes. To take a very simple example, suppose that, according to a certain theory of attitude change, changes in utility but not changes in belief can cause preference reversals of a certain type; if sour grape phenomena involve preference reversals of this type, then one can conclude that, according to these theories, sour grapes involves utility change. To apply this method, one needs firstly to identify some properties of the sour grapes phenomena, and then consider how different analyses of the change can account for these properties. In Section 2.1.2, some noteworthy properties will be identified; in Section 2.2, the possible analyses of the phenomenon will be explored. A decisive conclusion will not be reached; in Section 2.3, the direct method will be considered.

Remark 1 As indicated above, the difference between utility and expected utility turns out be very important to the understanding of the phenomenon of sour grapes. Given the preceding considerations, the important distinction is the following: utility is *pure* insofar as the calculation of the utility of a consequence does not depend on the beliefs of the agent, whereas expected utility is *mixed*, insofar as it depends not only on the agent's utilities for consequences, but also on his beliefs. However, this

[1] See Elster (1983, pp 112–114) for more details on why learning (a paradigmatic form of belief change) is not necessarily a problem for utilitarianism.

distinction is not always present in other frameworks, and, even where it can be drawn, it is often not explicitly recognised.[2]

First of all, the decision theory proposed by Jeffrey (1972), as well as the causal decision theories inspired by it (for example Joyce 1999), does not accept a distinction of this sort.[3] Whereas Savage distinguishes between the objects of pure utility (consequences), the objects of choice (acts), which have mixed utility, and the things over which the agent has no power (states of the world), Jeffrey employs a single ontology, consisting of "propositions" (sometimes also known as "prospects"). Each proposition has a probability and a utility (or "desirability") value, the latter being related to the probabilities and utilities of other propositions according to the following conditional expected utility formula: for incompatible propositions X and Y, $u(X \vee Y) = u(X).p(X/X \vee Y) + u(Y).p(Y/X \vee Y)$. So everything – from the choice of trying to get the grapes to the prospect of obtaining the grapes to the fact that they are 2 m off the ground – has a utility, and each of these utilities are mixed in the sense that they can be calculated from the utilities of other propositions. (Because of the technical conditions required, every proposition can be written as the disjunction of two incompatible non trivial propositions.) The general question posed by sour grapes – what changes: beliefs or (just) utilities? – makes no sense in this framework, for there is, in general, no change of utilities without a change in beliefs.[4]

Nevertheless, it does make sense to ask about the changes of utilities for specific propositions, or changes of beliefs in particular propositions. In the case of the fox, is it the utilities of the grapes which have changed, or the beliefs about their position? In the first case, there are beliefs which change (for example, beliefs as to whether he will get the grapes), in the second case, there are utilities which change (for example, the utility of the position of the grapes), but these do not seem pertinent for the problem posed by sour grapes. A simple way to understand the discussion of the majority of this paper, which will employ the Savage framework, would be to "embed" the Savage-style decision situations described above into the Jeffrey framework, and understand talk of 'utility change' and 'belief change' as referring to those utilities and beliefs which are present in the Savage framework (i.e. utilities for Savage consequences, and beliefs in Savage states). In this way, a meaningful distinction between pure and mixed utility change could be drawn in the Jeffrey framework.

[2] In cases where theorists of 'preference change' do not draw a difference between utility and preference or expected utility, their theories shall, insofar as is possible, be interpreted as theories of utility change rather than expected utility change; this interpretation, although debatable, at least allows us to consider what their theories can bring to the problem of sour grapes.

[3] Below, when treating alternatives to Savage, we will restrict attention to Jeffrey's evidential decision theory and not consider the causal decision-theoretic variant in any detail. Although there may be differences in the details and debate about which of the two is more naturally applicable to particular cases, the general points made here apply to both sorts of theory. Indeed, our intention is to avoid such debates and details, which are not central to the problem in hand.

[4] This is an immediate consequence of the fact that, in Jeffrey's decision theory, desirabilities determine probabilities (Jeffrey 1972, §5.9).

There appears to be no clear consensus as to what such an embedding should be.[5] Although the details are not important for the purposes of this paper, we shall take the following "brute-force" embedding.[6] For Savage states $s_1, s_2 \ldots$, let there be propositions $S_1, S_2 \ldots$ stating that the appropriate state is realised, and for Savage consequences $c_1, \ldots$, let there be propositions $C_1, \ldots$ stating that the appropriate consequence is realised (note that the S_i and the C_j are each sets of mutually incompatible propositions partitioning logical space). Let F be the proposition that act f is carried out. It follows from Jeffrey's conditional expected utility formula that $u(F) = \sum_{S_i} \sum_{C_j} u(C_j \wedge F).p(C_j/F \wedge S_i).p(S_i/F)$. If one assumes that probabilities of states are act-independent ($p(S_i/F) = p(S_i)$), that utilities of consequences are act-independent ($u(C_j \wedge F) = u(C_j)$), and that the probability of a consequence given an act and a state is one if the act sends that state to that consequence and zero otherwise ($p(C_j/F \wedge S_i) = 1$ if $f(s_i) = c_j$ and 0 otherwise), then one obtains the traditional Savage expected utility formula. Whatever is said about the Savage framework holds for the aforementioned propositions under the assumptions just stated; we shall consider these propositions and these assumptions to constitute the embedding.

The introduction of the distinction between pure and mixed utilities also requires a re-appraisal of several purported theories of "utility change" or "preference change", insofar as they can be applied to sour grapes. There are theories, both in the decision-theoretic and logical literature, which model "utility" or "preferences" as depending on beliefs and which account for utility change in terms of changes of beliefs: Cyert and DeGroot (1975); de Jongh and Liu (2006) are examples. They should be considered as models of change in *expected utility* (or, at least, change in mixed utility), and not of change in (pure) utility. If the fox turns away, as in La Fontaine's version of the fable, saying to himself that the grapes are sour, this may be understood as a change in belief about the taste of the grapes. The utilities for the grapes are thus to be understood as *mixed* utilities, which are a function of the (pure) utility of grape–properties and of the fox's beliefs as to whether these grapes have those properties have changed. Sour grapes will thus be analysed as a change in the fox's beliefs about whether the grapes have the grape–properties in question; under this analysis, the *mixed* utilities have changed, but the *pure* utilities have remained constant (he still enjoys the grape–properties as before). As noted above, the

[5] Normally in the Jeffrey framework, one explicitly does away *either* with states of the world *or* with consequences when considering the utility of an act. In Jeffrey's introduction to his theory (1972, Ch. 1), he does away with consequences, and considers explicitly only states of the world (which he calls "conditions"): using the terminology introduced in the text, $u(F) = \sum_{S_i} u(S_i \wedge F)p(S_i/F)$. Here the consequences are not explicitly represented but replaced by the conjunction of the realisation of the state and the act. By contrast, others (for example Levi (2000)) remove the states and leave the consequences: this yields $u(F) = \sum_{C_j} u(C_j \wedge F)p(C_j/F)$. Here the states are not explicitly represented, but are tied up in the conditional probabilities. Both of these can be derived from the general formula stated in the text: the former by summing over the consequences, the latter by summing over the states. See also Jeffrey (1972, §10.4).

[6] For a more sophisticated account of the relationship between the two frameworks, see Bradley (2007b).

intuition that the pure utilities as well as the mixed utilities change in sour grapes is important for the debate on utilitarianism (Elster 1983, p123); these theories deny this intuition in their analysis of the sour grape phenomenon.

2.1.2 Some Properties of Sour Grapes

Here are some observations concerning sour grapes inspired from Elster's classic discussion of the phenomenon.

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.

The source of the change Regarding the source of such changes, as Elster (1983, pp121–122) points out, one might draw a distinction between those caused by changes in the world – the "state-dependence" of preferences – and those caused by changes in the options open – "possibility dependence". 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).

To avoid confusion, it should be emphasised that the sort of dependence to which Elster seems to be alluding, especially in the case of state-dependence, is *diachronic* dependence. The *previous* states, choices and so on have an influence on the *current* utility. This notion is thus crucially different from the "state-dependence of utility" studied by decision theorists such as Karni and Drèze. A state-dependent utility is a utility that is not only a function of the consequence, but equally of the state which the act will "take" to that consequence. (The expected utility is calculated by the formula $\sum_{s \in S} p(s)u(s, f(s))$, with u having two arguments.) Note that the dependence in this case is entirely *synchronic*: in a single decision situation, at a single instant, the utility depends on the states. Although Elster only claims diachronic state-dependence for sour grapes, the synchronic or decision-theoretic sense is mentioned because it will be relevant below.

The permanence of 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 attitudes (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 individuals' attitudes seem to change "for good"; such is the case for the change of preferences for city or countryside life considered by Elster (1983, p112 *sq*.). 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. In Elster's example, someone who moves to the city may acquire a taste for city life, which may be reversed if he moves back to the countryside for a considerable period. A full analysis of sour grapes should be able to account for these factors.

2.2 Three Analyses

This section contains three analyses of sour grapes. In fact, to the extent that they leave precisely specified blanks to be filled by particular mechanisms, they are better described as analysis schemata. For each analysis, its capacity to account for the properties of sour grapes discussed above will be considered. This will not only allow us to ascertain whether the indirect method described in Section 2.1.1 can yield a reply to the question of sour grapes, but it will highlight some challenges for current theories of attitude change.

Throughout the section, the standard sour grape story introduced in Section 2.1 will be the principal object of consideration. Recall that this story involves two decision 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 the decision to try again or to give in (the 'second situation'). These situations seem to share the same 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. They also share the same set of consequences, which contains two elements – obtaining the grapes and not obtaining the grapes. It follows that the two situations share the same set of acts.

2.2.1 Pure Utility Change

The analysis 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 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 act f is $\sum_{s\in S} p_1(s)u_1(f(s))$. Then, according to this model of the change, 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 $\sum_{s\in S} p_1(s)u_2(f(s))$. A full model results when one adds an account of the change from u_1 to u_2; one might expect theories of utility change for example to provide such accounts.

Properties of the analysis

What has changed 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, in the following sense: a given change in utility will result in a change in preferences which is such that it there is *no* change in beliefs alone which could have produced this change in preferences. That is to say, one could not even rewrite the representation in the second situation $\sum_{s\in S} p_1(s)u_2(f(s))$ *as if* it consisted in a change of belief with a fixed utility (i.e. in the form $\sum_{s\in S} p'(s)u_1(f(s))$), because for these two sums to represent the same preferences, the probabilities and utilities must be the same ($p_1 = p'$ and $u_2 = u_1$).[7] One can thus conclude that this model of sour grapes essentially involves a change in utility.

The source of the change The source of the utility change is the fox's choice in the first situation and his experience following it: in this sense, the change is state-dependent. On the other hand, the same options (acts) are available in the first and the second situation, so the change cannot be thought of as possibility-dependent. Indeed, many of the current theories of preference or utility change keep the same options (i.e. possibilities) but alter the preferences on them[8]; to this extent, they could only be understood in terms of state-dependent change. As such, they fail to account for the tight relationship between state- and possibility-dependent change.

The nature of the change Theories of utility change are often motivated by examples involving changes in the face of statements specifying particular preference

---

[7] Suppose not, and, supposing appropriate calibration of the utilities, let p' be a probability such that $\sum_{s\in S} p_1(s).u_2(f(s)) = \sum_{s\in S} p'(s).u_1(f(s))$ for all acts f. Given an appropriately rich set of acts, as implied by the Savage axioms, for example, it follows that $p_1(s).u_2(c) = p'(s).u_1(c)$ for all states s and consequences c. However, given that s and c are independent variables, this implies that $p' = p_1$ and $u_1 = u_2$.

[8] Cyert and DeGroot (1975), van Benthem and Liu (2007), Bradley (2007a) and the revision and contraction operations in Hansson (1995) are examples.

relations to be accepted,[9] and it is not evident how to translate the changes in the world involved in this example – the fox's experience of his first attempt at getting the grapes, say – in terms of such statements. This just flags a general and important issue for theories of change: how is the trigger for a given change to be represented in the model? This question will always need to be posed when, as is the case here, theories are motivated with examples of *intentional* preference change (considered to be the preference analogues of apparently intentional processes such as learning from observation or accepting an announcement): something should be said about how one can apply the same methods for non-intentional changes (as noted in Section 2.1.2, in the case of sour grapes, the fox's experience *causes* the change, he does not *decide* to change his utility after the experience).

The permanence of the change The analysis does seem to account for the case where the subject actually acquires the preference in the long term, modelling it as a straight utility change. However, the modalities of the change seem to have been reversed: an adaptation of the preferences over a long period is captured here by a sudden revision of the utility at a particular moment. Furthermore, it is not certain to what extent 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. As for the case of the analysis in terms of belief change proposed below (Section 2.2.2), the capacity to account for such phenomena may depend on the particular theory of utility change adopted.

Many theories of pure utility change are modelled on, or related to, particular theories of belief change. The theory of Cyert and DeGroot (1975) is not strictly speaking a theory of pure utility change but of mixed utility change, insofar as utilities depend on beliefs and the change in utility arises from a change in beliefs, notably by Bayesian conditionalisation. *Even if* it were possible to understand this as a theory of pure utility change, it would inherit the reversibility properties of Bayesian conditionalisation, and notably the fact that this sort of change is generally irreversible – information is lost in the change. Similarly, approaches modelled on public announcement logics also tend to yield change operators which are irreversible (the update and upgrade operators in van Benthem and Liu [2007] seem to be examples).

On the other hand, some theories which may be understood as theories of utility change are inspired by the literature on belief revision: such is the theory of Hansson (1995). They have two operations: revision, which establishes an order on the utilities of various consequences, and contraction, whereby one revokes a particular order on the utilities of consequences. An application of the former operation followed by an application of the latter (with respect to the same order on the same consequences) yields the original utility: in this sense, there is reversibility. Similarly, the theory of Bradley (2007a) according to which utility values are re-allocated

[9] 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.

on events in a partition, is reversible: it suffices to apply a change operation allocating the original utility values to the elements of the partition to return to the intial utility function.

Although technically, so to speak, these theories give reversibility, the sorts of concerns mooted above regarding the representation of the trigger for change apply here: if, as the fox walks away, he sees a ladder and uses it to get the grapes, is this correctly understood as a retraction of the original utility change whereby his utility for the grapes decreased (as in the case of Hansson's theory) or as the demand to acquire utilities for grapes which happen to be those he had initially (as in the case of Bradley's theory)? Even if this is a correct analysis of the case of short-term reversibility, will the same mechanisms be at work in the long-term case (when he reverts backs to a high utility for grapes after years of having a low utility for grapes)? A final, general worry concerns the intuitive adequacy of any analysis of short-term reversibility of preferences in terms of a succession of utility changes: is it really plausible that the fox's utility undergoes two abrupt changes in such a short period of time? The reversibility and permanence phenomena pose interesting challenges for anyone seeking to take up and defend this analysis, and it is not clear that they are fully met by current models of utility change.

This first analysis of sour grapes is firmly embedded in a developing theory of utility change. At this stage, all that can be noted is the difficulties which should overcome: it does not seem to capture the subtle relationship between the state- and possibility-dependance of the change; careful interpretation of the model is required to make sure it can cope with the non-intentional nature of the change; and there are doubts regarding its capacity to account for the permanence and reversibility properties of the change. It is debatable how many of these difficulties support the conclusion that sour grapes is not a case of utility change and how many are to be seen as concerns with the adequacy of current models of utility change.

Nevertheless, even under this meagre construal, some may find the analysis in terms of utility change inadequate. There is an intuition, which seems incompatible with this sort of analysis, that the fox's utilities do not *really* change, at least not immediately after his failed attempt at getting the grapes. The second analysis of sour grapes takes this as its guiding intuition.

2.2.2 Self-justification

The analysis An important intuition about the sour grapes phenomenon is that it does not involve so much the *action* of the agent (at the moment of the sour grapes phenomenon) as the *way he justifies* or *rationalises* that 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) affect concurrent behaviour, the rationalisation he constructs for himself will affect the utilities and the beliefs *he sees himself as having*.

In this analysis, crucial use is made of the distinction between the point of view of the modelee – the agent – and that of the modeler – the decision theorist. 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. This allows one to distinguish between an *internal* model, representing the utilities and the beliefs the fox sees himself as having, and the representation of a competent external observer. According to the analysis 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, as opposed to the case discussed in Section 2.2.1, a change in beliefs always produces a change in preferences (expected utility) which can also be produced by a change in utilities (with constant beliefs). This is the change that the fox considers to have occurred: he represents the change to himself as a change in degree to which he values the grapes, that is, as a change in 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 $\sum_{s \in S} p_1(s)u_1(f(s))$. 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 initial probability and utility functions: that is, they represent the fox's attitudes in the first situation, *according to both the fox and the modeler*. The modeler and the fox will disagree however on the representation of the fox's attitudes 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). After the change, the fox's expected utility thus becomes $\sum_{s \in S} p_2(s)u_1(f(s))$. 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 $u_2(s, c) = \frac{p_2(s)}{p_1(s)}u_1(c)$. (Note that u_2 is a state-dependent utility: it is a function of states and consequences; see Section 2.1.2.) This is the sort of change that the fox *sees himself* as undergoing: according to him, he has not learnt that the grapes are harder to obtain, he has just changed his mind about whether he wants them or not. As for the analysis in Section 2.2.1, this is only a general schema: different concrete models are obtained by adding particular theories of belief change.

Properties of the analysis

What has changed Concerning the question of what has changed, the point of view taken on the situation is crucial. All are agreed that the expected utility 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 attitudes one considers oneself to have.

The nature of the change There are two aspects of the change (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. Hence interpretations of belief change mechanisms as models of "unintentional" change are pertinent here; mechanisms which support such interpretations apply more naturally.

The source of the change The role of the first attempt, and more particularly the influence of past states and choices on the preferences in the second situation, indicates that there is (diachronic) state-dependence. Moreover, the fact that the same states and consequences are involved, and the same acts are on offer, in the first and second situations implies that this analysis does not consider the changes as possibility-dependent. It thus cannot account for the subtle relationship between state- and possibility-dependence.

The permanence of the change From one point of view, this analysis seems amenable to reversibility: since the utility of the fox "really" remains the same (from the modeler's point of view), it is no surprise if he "reverts back" to this utility. However, according to this analysis, the changes in preference are due to changes in belief; therefore it is the theories of belief change which will have to explain the observed changes in preference, and notably the reversibility phenomenon. The capacity of theories of belief change to account for the reversibility may depend on the theory considered.

According to many major theories of belief change, changes in belief are not in general reversible; thus the short- and long-term reversibility may not be accounted for by versions of this analysis which use such theories.[10] This is the case for the most important quantitative model of belief change, Bayesian conditionalisation: under conditionalisation by an event, information about the probabilities of events which are incompatible with the conditionalising event is lost. Similarly points can be made for logical theories of change whose operators are analogous to conditionnalisation: public announcement logic is an example (van Ditmarsch et al. 2007). Although such theories do not easily account for reversibility, they do naturally capture the long-term change in preferences, in terms of the long-term changes of beliefs brought about by, say, conditionalisation.

On the other hand, reversibility is allowed by the generalisation of Bayesian conditionalisation proposed by Jeffrey (1972): one revises by modifying the

[10] This indeed is in harmony with Elster's use of the reversibility phenomenon to distinguish sour grapes from learning (Elster 1983, pp112–114).

probabilities of events in a partition, without generally setting any of the probabilities to zero or one, so that one can reverse the change by setting the probabilities back to their initial values. Similarly, traditional AGM theories of belief revision (Gärdenfors 1988) exhibit reversability insofar as the operation of expansion, by which new information consistent with current beliefs is incorporated, followed by the operation of contraction, by which beliefs are removed from the agent's corpus, yields the initial set of beliefs.[11]

Points similar to those made in Section 2.2.1 concerning the proper representation of the trigger for change in the formal model hold here: if, immediately following the fox's decision to walk away, there arose an opportunity to get the grapes (he spots a ladder, for example), is this to be thought of as a retraction of some belief he had acquired about the states of the world (a contraction) or the acquiring of the new information that the probabilities of the states of world were as he had originally thought (as in the case of Jeffrey conditionalisation)?

This analysis has the advantage over the previous one that there is a larger amount of work on belief change to draw upon. Furthermore, whereas several belief change operations cannot adequately capture some of the subtle properties of sour grapes, in particular relating to reversibility, others seem more capable. Nevertheless, the analysis does share some of the potential difficulties of the previous one: the relation between state- and possibility-dependance is not accounted for, and subtle interpretation of the belief change operations is required to account for the non-intentional nature of the change. Once again, it is unclear whether these are to be taken as indications that sour grapes is not a case of belief change, or simply as a challenge to be overcome by theories of belief change.

Moreover, there are several other aspects of the model which some might find unsettling. There is a certain intuition according to which there is no change of beliefs involved in sour grapes: the beliefs of the fox in this model concern the position of the grapes, the height of the fox and similar information, and he knew all of this information *before* his first attempt. So what has he learnt? The natural answer seems to be that he has learnt the chances of success at obtaining the grapes, given that they are at such a height. As just noted, these do not correspond to beliefs represented by his probabilities over states of the world. Rather they correspond to a fact about the decision situation he is in: namely, to how reliably the acts on offer effectuate the transitions from states to the consequences which they claim to. This is a guiding intuition for the third analysis.

2.2.3 Reliability of Acts

The analysis 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 for

[11] It seems that the same cannot be said of dynamic logic approaches to belief revision, such as that developed by Baltag and Smets (2006), for they do not have a contraction operation.

consequences, but he *alters the way he represents the decision problem he is faced with*. A concise way of representing this change is by the addition of a *situation-dependent* or *context-dependent* factor in his calculation of expected utility. This factor can be left explicit, where it receives a natural interpretation as a change in his opinion about the reliability of the acts on offer; on the other hand, it can be absorbed into the utility function and thus be interpreted as a change in it.

Formally, the simplest proposal is to introduce a real-valued function on the pairs of states and consequences (s, c), call it γ, which features in the representation of preferences in the second situation. For initial probabilities and utilities p_1 and u_1, the preferences in the first situation are represented by the ordinary expected utility formula, whereas the preferences in the second situation are represented by $\sum_{s \in S} p_1(s).\gamma(s, f(s)).u_1(f(s))$. Just as the other sections presented analysis schemata to be filled in with theories of utility change and belief change respectively, this is but a type of analysis: a concrete analysis is attained by adding a theory of the factor γ.

There are several possible interpretations one could give of this factor. A natural one was mooted above: γ is a *reliability factor*, which reflects the chances of success of an act purporting to take a given state to a particular consequence. If $\gamma(s, c) = 1$, an act purporting to take state s to consequence c will certainly succeed, if $\gamma(s, c) = 0$ it will certainly fail and for intermediate values it will have intermediate chances of success (assuming γ to be normalised to take values in [0, 1]). As such, γ can be thought of representing *constraints* on whether acts can deliver particular consequences given particular states: it is thus something which is built into the agent's representation of the decision problem he is facing.[12] This is the principal difference with respect to the analyses of sour grapes proposed above. As remarked in Section 2.1.1, the agent separates the possible consequences of his actions, the properties of the world which determine whether these consequences are reached, and considers the options to be functions taking these possibilities to the appropriate consequences. Under the first analysis, it is the utilities of the consequences which change after the fox's first attempt at getting the grapes. Under the second analysis, it is the beliefs about the world. Under the current analysis, it is the very structure of the decision problem which the agent is assuming that changes. The factor γ is thus properly thought of as a *context-dependent* or *situation-dependent* factor which reflects aspects of the decision problem; the sort of change currently under consideration is a change in the agent's perception or representation of the choice he is facing.

Remark 2 Naturally, the way the agent represents the decision problem encapsulates the things which he presupposes to be true of that problem. To the extent that these *presuppositions* can be qualified as beliefs, the change can be thought of as a sort of change of belief (for a model of presuppositions, and a discussion of its relation to explicit beliefs, see Hill [2008b]). However, as noted above, this change in

[12] Note that it is traditional practice in economics to build the appropriate constraints into the decision problem.

belief is not a change in his beliefs about the states of the world (which in this case specify the position of the grapes) and so this is not a change in his probabilities, as was the change discussed in Section 2.2.2. Of course, the line between explicit beliefs and presuppositions varies with the representation of the problem. It is possible to represent the decision problem so that the issue of the reliability of acts is left open: it suffices to use states of the world which specify the reliability of the acts carried out in them (for example, in a given state of the world, the grapes are at a height of 2 m and the act of grasping them is reliable, in another state, the grapes are at 2 m, and the act is not reliable, and so on). If the decision problems for the fox's first and second attempt are represented like this, the presupposition about the reliability of acts becomes an explicit belief, represented by probabilities, and the change is of the same sort as in Section 2.2.2. However, the sorts of change which are of interest here, and which differ from those discussed in the previous sections, are changes in the representation of the decision problem (or, in other words, changes in the presuppositions). The use of the factor γ is merely one way of representing such a change. It is, however, not the only one.

Another possibility is to represent the change as involving a change in the set of states – and therefore the set of acts – in the representation of the decision problem and the introduction of a probability function on the new set of states which is suitably related to the probability function on the old set of states (in particular, such that they agree on common events). For example, the set of states used up to now in this section will be replaced by states mentioned in the previous paragraph which specify not only the position of the grapes but also the reliabilities of acts. Models of such a change can be supplied by models of states of belief which can cope with awareness and awareness change (for example, Hill [2007b, 2008a], or in a probabilistic setup Modica [2008]).

A third possibility is afforded by the Jeffrey framework. (As elsewhere in the paper, we concentrate attention on Jeffrey's evidential decision theory, without taking sides in the debate which opposes it to causal decision-theoretic variants.) Recall that, under the assumptions that the probabilities of states are act-independent and that the utilities of consequences are act-independent, we have $u(F) = \sum_{S_i} \sum_{C_j} u(C_j) p(S_i) p(C_j/F \wedge S_i)$ (see Remark 1 for discussion and notation).[13] Recall that one obtains Savage's expected utility formula under the assumption that the third term in the product is equal to one when $f(s_i) = c_j$ and is zero otherwise. Weakening this assumption introduces a third term into the traditional expected utility formula and is thus reminiscient of the introduction of the factor γ proposed above. Hence, in the Jeffrey framework, the change in preferences can be represented by the fact that this third term ceases to take the standard values assumed in the traditional expected utility formula. In other words, the change is a change in the probability of reaching the consequence given the state and the act which purports

[13] It was noted in footnote 5 that in the Jeffrey framework it is natural to remove the states, by collecting the $p(S_i)p(C_j/F \wedge S_i)$ terms into a single $p(C_j/F)$ term. Doing this removes the probabilities of states of the world and hence obscures the distinction between the analysis discussed here and that of the previous section. For this reason we do not consider this option further here.

to yield that consequence on that state. Naturally, this is very close to the reliability interpretation of the factor γ. Furthermore, just as for the proposals in the previous paragraphs, under this analysis the change does consist in a change of belief, but not a belief about the states of the world: it is rather a belief about the decision situation which the agent is faced with, and more particularly about the real effects of the options on offer. If the change is understood this way, theories of change in these conditional beliefs, such as that proposed by Bradley (2005), can be used.

Our purpose here is to draw attention to the presuppositions or factors which reflect the decision situation the agent considers himself to be in and to the fact that they change and could be at the root of some preference changes. We do not claim that the above considerations provide anything close to a comprehensive theory of the relevant aspects of the decision problem, nor a comparison of possible representations of them. Indeed, we do not even claim that the three proposals discussed briefly above are equivalent. It is evident that they are not: for example, whereas the first proposal always yields preferences which satisfy Savage's most fundamental axiom on preferences – the sure-thing principle – the last one does not.[14] Indeed, in the light of differences such as these, it may be that the most appropriate representation of the situation-dependent factor will depend on the framework which one is using (see also Remark 3, Section 2.3.2). Given that the framework adopted here is Savage's, and the factor γ seems to be the option which fits in easiest and most simply with this framework, we shall focus on this option for the rest of the paper. The discussion is intended to carry over to other representations of the situation-dependent factor, although it remains to be seen on a case by case basis to what extent it does.

Properties of the analysis

What has changed Note firstly that, like the first analysis (Section 2.2.1), it is not always possible to reformulate the change in the fox's preferences as if it consisted in a change in belief regarding the states of the world.[15] However, it *is* always possible to reformulate it as if it were a change in utilities: $\gamma(s, c).u_1(c)$ may be thought of as a utility function. (Note this utility function is state-dependent.) In this sense, the experience can be thought of as causing a *change in the utility function* from the initial utility u_1 to $\gamma(s, c).u_1(c)$; what is more, γ characterises exactly this change. Evidently, it is *not* necessary to see this as a utility change, because $\gamma(s, c).u_1(c)$

[14] For the former claim, see Section 2.3. As for the latter claim, given the presence of non-zero terms $p(C_j/F \wedge S_i)$ where $f(s_i) \neq c_j$, there will be utility contributions from pairs s_i, c_j where $c_j \neq f(s_i)$, and this contradicts the sure-thing principle. Changes in the set of states may also produce violations of the sure-thing principle with respect to the initial states, if one translates the acts on the initial set of states to acts on the new set of states in the appropriate way; see Savage (1954, §5.5) and Hill (forthcoming, §4).

[15] 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. See also Remark 2.

is not the only utility involved in second situation: there is still the initial utility u_1. This is, so to speak, the *absolute* utility, independent of the situation, whereas $\gamma(s,c).u_1(c)$ is the utility *in this situation* – relative to the situation insofar as the situation limits, through the factor γ, the accessibility of the consequence c, or the chances of actually obtaining this consequence.[16]

Several of the attractive properties of the analysis come from the distinction between the utility "effective" in the decision and the agent's "underlying" utility. The first of these properties has already been evoked: in this analysis, sour grapes comes out as a *change in the situation-relative utility*, though not in the *absolute utility*.[17] This has interesting consequences for utilitarianism: should one use absolute utilities in considerations of social good or more variable situation-relative utilities? The argument against utilitarianism mentioned at the beginning of the paper is evidently stronger in the latter case. Moreover, the distinction between the utility which the agent effectively employs and some more stable underlying utility is reminiscent of other distinctions which have been recently proposed in decision theory, such as the distinction between "experienced" and "intrinsic" utility argued for by Kahneman et al. (1997). Naturally, one might either interpret the fox as not being conscious of the duality of utilities, and thus as thinking that his utility has changed, or as being lucid as to the pair of utilities, and thus aware, when he mumbles that the grapes are no longer desired, that his affirmation is context-relative and refers to the situation-relative utility, not the absolute utility.

The nature of the change As noted above, the change is a change in the representation of the decision problem which the agent is faced with. Some of the models for changes of this sort only admit natural interpretations according to which the change is non-intentional. Above it was noted that the change could be captured by a change in the set of states: this would involve an increase in the awareness of the agent, and one cannot intentionally decide to become aware of something (Hill 2007b). Similarly, there is a sense in which it sounds strange to intentionally question a presupposition (one is usually "forced" or "led" to question it). On the other hand, some models support intentional as well as non-intentional readings, and the latter are required here. For example, it was noted that, in the appropriate framework, the change can be thought of as a change in certain types of conditional beliefs; as for the literature on belief and utility change (Sections 2.2.1 and 2.2.2),

[16] To an extent, this duality in utilities can be reproduced under the other two proposals mentioned in Remark 2. In the Jeffrey framework, the situation-relative utility is $u(C_j \wedge F)p(C_j/F \wedge S_i)$; note that, unlike the proposal in the Savage framework, this utility is not only state-dependent, but also act-dependent. As for the change in the set of states, such a change is equivalent to a change in the set of consequences (see the "small world consequences" in Savage [1954, §5.5]); the utility for these new consequences can be thought of as the situation-relative utility.

[17] The interpretation of the difference of these utilities is not unrelated to the interpretation of the factor γ. The fact that, in general, γ is a way of capturing an aspect of the decision situation the agent sees himself as faced with justifies the terminology 'situation-relative utility'. In the particular case that γ is considered to reflect the reliability of acts, one could give this utility another name, such as 'reliability-discounted utility'. For other interpretations of the difference between the two sorts of utility, see Hill (forthcoming).

most of the discussions of such changes consider examples where the incoming information is in the form of new conditional beliefs which are to be intentionally incorporated into one's corpus (Bradley 2005).

The source of the change In understanding the factor γ as encapsulating an element presumed to be integral to the decision problem the agent perceives, one can easily take account of the flexible distinction, noted in Section 2.1.2, between the state- and possibility-dependence of the change. On the one hand, the change between the two situations can be considered as a result of the choice in the first attempt and his experience of the results of that choice: there is thus diachronic state dependence. On the other hand, it was noted above that the factor γ represents constraints on the range of acts he can expect to carry out successfully. In other words, the introduction of this factor implies a change in the possibilities *effectively available* to the fox in second situation. In this sense, there is possibility-dependence.

The permanence of change The fact that the belief and absolute utility are left fixed, but that some sort of situation-dependent factor intervenes, allows a rather intuitive understanding of short-term reversibility. The choice of walking away from the grapes, and the choice of using the ladder which the fox notices as he walks away to get them are made in different situations, and it certainly seems that the chances of success of the act of attempting to get the grapes differ between these situations. These are two different decision problems, involving different situation-dependent factors. This is but an intuition: naturally, an adequate theory of the representation of the decision problem or of the situation-dependent factor, which accounts for such differences in an appropriate way, is required.

As for the long-term permanence of the preference change, this may perhaps be modelled by the fact that, throughout an extended period, there is a generally similar sort of decision problem, with a similar set of acts and a stable situation-dependent factor; once again, such a model remains to be developed formally. Such a model, combined with the duality of the utilities, would allow an understanding of the intuition that it is the agent's utility that changes over such extended periods, by reasoning as follows. Since the situation-dependent factor is generally stable over the period, the same situation-relative utility applies throughout the period. If one comes to presuppose on this basis that the situation-dependent factor is fixed at that value, the situation-relative utility, which integrates this factor, does not vary during the period and can be effectively treated as the "real" utility in decision problems during this period, usurping the absolute utility. It would thus seem that the intuition that a change of utilities is involved in such cases actually refers to the difference between the absolute utility (which is the same before and after the change in preferences) and the situation-relative utility (which appears to be constant because of the stability of the sort of decision problem faced). Given this account of the long-term permanence of the preference change, long-term reversal of preferences could be understood in terms of a change of circumstance which is drastic enough to invalidate the presupposition that the situation-dependent factor is fixed – that is, by a decision problem with a different situation-dependent factor – in such a way that resort to the pure utility is once again required.[18] Under this account, long-term reversibility is a similar sort of effect to short-term reversibility, though perhaps of differing degree.

Just as for the other analyses, this is more a sketch than a fully developed analysis, insofar as a fully worked-out account of the representation, appearance and dynamics of the situation-dependent factor is still lacking. Nevertheless, depending on one's model of this factor, this analysis sits more or less easily with theories of change which deal explicitly with cases of non-intentional change, it is able to account for the flexible distinction between state- and possibility-dependent factors and it does seem to be subtle enough to cope with the reversibility and permanence phenomena.

2.3 Getting Your Teeth into Sour Grapes

What changes in a case of sour grapes – utilities, beliefs, or some aspect of the decision problem as the agent perceives it? An indirect method for answering this question would identify properties of sour grapes which, according to theories of attitude change, only occur in, say, changes in utility. It is fair to say, on the basis of the considerations in the previous section, that this method does not yield any strong conclusion: many of the doubts which could be emitted with respect to the analyses, such as the question of the intentional nature of the changes and to a certain extent the reversibility phenomena, seem to be as much challenges for theories of belief or utility change as they are specific features of one or other sort of change. The apparently more substantive differences between the analyses, such as the ability of only the third analysis to capture the subtle relationship between state- and possibility-dependence, seem too slight to support a firm endorsement of one at the expense of the others. To answer the essential question of sour grapes, a more direct method is required.

Such a method cannot rely solely on theories of attitude change, because these theories generally assume that one knows the attitude which is to change, and go on to say how it should change; that is, they assume that the answer to the central question of sour grapes is already known. Something else is needed: a way of determining what the agent's attitudes are at any given moment. Given this, one could see what has changed by eliciting the agent's attitudes before and after the change and comparing them. Note that this method applies most naturally to *instances* of preference change, rather than to the *class* of changes which can be classified as cases of sour grapes: indeed, given that the properties of the sour grapes phenomenon are not sufficient to deduce what sort of change are occurring, it is possible that in different cases of sour grapes-style phenomena, different sorts of changes are taking place. Throughout this section, an arbitrary particular instance of preference change will be considered.

[18] Indeed, in Elster's example of long-term reversibility (of the preference for city life; see Section 2.1), a drastic change is required (a move to the countryside for a considerable period).

To the problem of change which has been considered up to now – the problem of understanding the relationship between the agent's attitudes *in different situations* – we thus add the problem of *elicitation* – that of distinguishing the part of the agent's preferences which is due to his beliefs from the part which is due to his utilities *in a given situation or decision problem*. To answer the question posed by sour grapes, one requires a reply to both problems at once: a way of distinguishing between the agent's beliefs and his utilities in a given situation *which is such that* it yields an understanding of the changes they undergo.

2.3.1 The Direct Method: The Classic Approach

Decision-theoretic preliminaries The question of elicitation has not been a subject for logicians, but for decision theorists. At the heart of every decision theory is a *representation theorem*, giving a set of necessary and sufficient conditions on the agent's preferences for there to exist a unique probability and an essentially unique utility representing the preferences (according to the expected utility formula, in the present case). The uniqueness of the probabilities and utilities is crucial: it allows proponents to claim that they represent the beliefs and desires underlying the agent's preferences. Related to these theorems, or their proofs, there are often methods for practically eliciting or approximating the agent's probability and utility functions from their choices. The conditions in the representation theorems become assumptions that are needed for the success of the elicitation techniques. (Beyond the practice of behavioural decision theorists, this sort of connection between norms on rational preference and the conditions for the possibility of understanding the agent by attributing attitudes to him is present in philosophy, notably in the work of Donald Davidson.) At least this is so with the most developed and simplest decision theory, namely the theory of expected utility pioneered by Savage, and we shall focus almost entirely on this theory for the rest of the paper.[19]

Representation theorems such as Savage's generally work with a single decision situation: a single set of states, consequences and acts, with a single preference relation over the acts. Although the axioms in the theorem are formulated as *conditions on the preference relation*, they correspond more or less neatly to *principles concerning the rationalisation of the agent's preferences*. In endorsing the conditions on preferences as norms on rational preferences, one is committed to endorsing the principles governing the rationalisation of his preferences, and vice versa. Given that the details of the axioms on preferences are not important for the purposes of this paper, we present the underlying principles involved in Savage's theorem, with only very rough indications as to the corresponding condition on preferences; for precise formulation and discussion, see Savage (1954).

---

[19] For some comments concerning the application to the Jeffrey framework, see Remark 3 below.

Order [P1] The preference relation is representatable by a real-valued function on the set of acts. (Condition on preferences: the relation 'is preferred to' is a transitive and complete order.)

Independence [P2, also called the Sure-Thing Principle] The aforementioned real-valued function on acts can be decomposed into the sum $\sum_{s \in S, c \in C} V(s, c)$ where V is a real-valued function on state-consequence pairs. (Condition on preferences: the preference relation over two acts depends only on the states where they differ.)

State-independence [P3 and P4] The function V can be decomposed into a probability p on states and a state-independent utility u on consequences. (Condition on preferences: for any pair of consequences, the preference order over them given any event is the same; for two-valued acts, the preference relation over them depends only on the set of states where they take the more preferred consequence, and not on the consequences themselves.)

Continuity [P6 and P7] The functions mentioned at the previous levels exist for large sets of states and are suitably unique. (Condition on preferences: the set of states and the preference relation are sufficiently rich and well-behaved under taking limits.)

For any agent satisfying these principles, and hence for any agent whose preferences satisfy the corresponding axioms, there exists a unique probability over states p and a suitably unique[20] state-independent utility u such that he prefers an act f to g if and only if the expected utility of the former is greater than that of the latter: $\sum_{s \in S} p(s).u(f(s)) \geq \sum_{s \in S} p(s).u(g(s))$.

A detailed discussion of these principles and the corresponding axioms on preferences is beyond the scope of the paper (see for example Savage (1954); Broome (1991)). Although there is reason to question each one, it is safe to say that the first two are often taken to have some normative justification and the last is largely considered to be "technical" or "structural". As for the second last (state-independence), it has been challenged by decision theorists such as Karni and Drèze, who have proposed alternative principles. It is worth underlining that none of these principles can be dropped without being replaced: each is necessary for the representation theorem and the possibility of eliciting attitudes. For example, the theorists mentioned above who challenge state-independence but retain the Savage framework have had to propose alternative, more complicated principles such that it is possible to elicit unique probabilities and state-dependent utilities from any agent who satisfies the new principles.

The common strategy and its weaknesses The natural strategy for determining what has changed in a particular instance of sour grapes employs theories of decision such as the one described above in the following way: assume that the agent's preferences satisfies the conditions for elicitation before the preference change, and use the elicitation methods mentioned above to elicit his probabilities and utilities;

[20] Precisely: unique up to positive affine transformation.

assume the agent's preferences satisfy the conditions after the preference change, and elicit his (new) probabilities and utilities; finally, compare the two sets of attitudes to see what has changed. We shall argue that traditional elicitation techniques which are more or less loosely related to representation theorems such as Savage's are not propiscious for use in such a strategy. The norms and conditions which guide the elicitation are entirely synchronic, and hence surrender any pretension of remaining faithful to features of the dynamics of the agent's attitudes.

The point can be made on the case of sour grapes. Suppose that the fox satisfies the principles above, and hence the corresponding preference axioms, in the first situation (when he decides to try to get the grapes) and in the second situation (when he walks away); hence the typical representation theorems of decision theory – Savage's theorem, for example – apply and his probabilities and utilities can be elicited. Suppose furthermore that this method of elicitation yields probability and utility functions according to which the first analysis (Section 2.2.1) is correct: the probabilities elicited in the two situations according to Savage's theorem are the same but the utilities have changed. Suppose finally that as the fox walks away from the grapes, he spots a ladder, seizes it, clambers up and grabs the grapes. This situation seems naturally understandable in terms of his high utility for the grapes (relative to not having them). Indeed, assuming that the agent satisfies the above principles (and hence the corresponding preference axioms) in this third situation, his attitudes can be elicited, and it indeed turns out that he has the same high utility for grapes which he had initially. Now we have a clash of intuitions. On the one hand, this sequence of situations and actions constitutes the (short-term) reversibility phenomenon recognised in Section 2.1.2; as noted in Section 2.2.1, it is unnatural to explain this phenomenon by a pair of sudden changes in utility. On the other hand, the strategy under consideration for deciding what has changed in an instance of sour grapes, which uses the repeated application of Savage's representation theorem and related elicitation techniques, implies that there is just this erratic pair of utility changes. What is going on?

Generalising from this example, consider two decision situations σ_1 and σ_2, perhaps with different acts on offer but with either a common set of possible consequences of these acts (so the states may be different) or a common set of relevant states of the world (so that the consequences may be different). The tension occurs when (1) there is a strong intuition that the agent's utilities (respectively beliefs) are the same in σ_1 and σ_2 but (2) he satisfies the conditions required in Savage's representation theorem in both σ_1 and σ_2, and the elicited utilities (resp. beliefs) differ. The previous paragraph contains an example of this sort for utilities (with σ_1 being the second situation in the story, and σ_2 the third); for a numerical example involving beliefs, see the interpretation Hill (forthcoming) offers of an example proposed by Karni (1996, pp 256–257). In examples such as these, one could ignore the intuition about the stability of attitudes, and follow the results of the elicitation blindly. However, the tension seems to indicate that there is an element of rational behaviour which is not captured by the principles underlying results such as Savage's. Namely, a rational agent's attitudes should not change gratuitously between appropriately related decision situations. Let us call this the *stability principle*.

Stability In the absence of a reason for the agent's probability or utility to differ between suitably related decision situations, they remain the same.

Naturally, like our rendition of the principles underlying Savage's theorem, this principle is formulated in terms of the rationalisation of the agent's preferences, rather than as a direct condition on the preference relation. As discussed briefly below, this principle can be translated into precise (behavioural) conditions on preferences, which feature in a representation theorem and serve as necessary and sufficient conditions for the success of an elicitation procedure. Under the translation, the concepts featuring in the principle as stated above are sharpened. For example, the conditions on preferences imply that the set of suitably related decision situations between which there is no reason for the agent's probability or utility to differ has certain properties; as such, they will provide a minimal axiomatic characterisation of this notion. Of course, in practice, it is up to the good judgement of the elicitor to decide what counts as an appropriate related situation, just as it is up to him to provide an adequate representation of the decision problem (the sets of states and consequences, in the case of Savage). Further discussion of the notion of suitably related decision situation is beyond the scope of this paper; the reader is referred to the discussion in Hill (forthcoming).

The examples considered above show that stability and the four aforementioned principles of classic Savagean decision theory may enter into conflict. In hindsight, this is not at all surprising. The Savagean axioms, like all axioms in decision theory, are synchronic: they deal, at least initially, with attitudes and choices in a given, fixed decision situation. By contrast, the stability principle is diachronic: it explicitly takes into account the relationship between decision situations. One cannot expect synchronic principles to be able to account for diachronic properties; indeed, the methods built on the Savagean principles fail to yield an intuitive understanding of the diachronic behaviour of the agent's beliefs and utilities in some cases. Do we have reasons for accepting the diachronic stability principle? And what is the price of accepting it?

Why stability? The first argument in favour of the stability principle comes from our intuitions. How do you tell whether the fox has *really* changed his mind about the desirability of the grapes? 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). Such folk wisdom invokes the stability principle in the two senses mentioned above. Firstly, as a norm: a rational agent should not change his utilities between such appropriately related situations. Secondly, as a guide to the agent's attitudes, and thus a way of understanding his behaviour: to elicit his utilities as he walks away, it is sufficient to elicit them in the related situation where, at just that moment, he spots a ladder. We do seem to use the stability principle in our rationalisations of human behaviour.

A second consideration in favour of the stability principle relates to the general project of understanding change. 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" on which theories of change build. Without the stability principle, attitudes risk being too erratic to allow such a null state. This was seen above in examples where, using elicitation methods which ignore inter-situational comparisons, utilities differ between situations where there is supposed to be stability. With such chaotic behaviour of attitudes between situations, meaningful investigation of attitude change becomes well-nigh impossible. If the problem of change is to be treated as well as that of elicitation, the principles underlying the understanding of an agent's behaviour and the elicitation of his attitudes must accommodate the basic requirements for an investigation into change. The stability principle is a way of doing this.

Regarding the price to pay for acceptance of this principle, note that the stability principle does not contradict the four standard Savagean principles. It may well be that the agent satisfies the Savagean principles in all decision situations, and that the probabilities and utilities elicited using standard techniques are the same in suitably related decision situations. However, the examples above seem to indicate that there may be cases where the Savagean principles and stability cannot be simultaneously respected. If the stability principle is to be retained as a norm for rational choice, then one of the Savagean principles will need to be sacrificed. Which one?

Stability only involves the probabilities and utilities of the agent. Hence it does not interfere with the order condition or the independence condition, which feature in the representation theorem, and implicitly in the process of elicitation at a stage before probabilities and utilities have been separated out. The two most important principles of Savagean decision theory are thus fully retained on adoption of the stability principle. Putting aside continuity, which is technical and does not explicitly involve the probabilities and utilities, all that remains is state-independence; and indeed, this is the condition which traditionally allows one to separate the probability from the utility. If stability is adopted as a guiding condition for elicitation of attitudes, then it is in place of state-independence. However, the normative and descriptive validity of the axioms for state-independence, as well as the principle stating that the agent's utilities are always state-independent, have been doubted, both by economists (Drèze 1987; Karni and Mongin 2000; Arrow 1974) and philosophers (Joyce 1999), so much so that it is fair to say that there is a consensus that, although they hold in some decision situations, they surely do not hold in others. Indeed, for many of these authors, they do not in general constitute an acceptable norm for rational decision, and thus cannot be assumed in the elicitation of attitudes. Above, when arguing that traditional methods are inadequate for the elicitation of attitudes in cases of change, it was assumed that all the Savage axioms held in all situations, so that the traditional techniques always work. However, the situation is in fact worse for the defender of traditional representation theorems and elicitation methods: in some situations, the conditions on preferences do not hold and the techniques cannot even be applied. The stability principle replaces the weakest plank in the Savage construction.

2.3.2 An Alternative Approach

Using stability Stability says that the agent's utilities (respectively probabilities) are the same in appropriately related decision situations. This allows a simplification in one's elicitation of the agent's attitudes. To elicit his utilities in the current decision situation, one can elicit his utilities in any appropriately related situation: assuming stability, the result will be the same. So choose the situations which are the easiest for elicitation: not only will the use of these situations make the elicitation easier, but by consequence it will be rendered more reliable.

Consider once again the sour grapes example, and consider the task of eliciting his utilities as he walks away from the tree. This is a complicated case for the elicitor: the fact that he does not choose to attempt to get the grapes is not an indication that he does not value them, because his beliefs may play an important role. It is much easier to elicit his attitudes in a situation where the grapes are offered on a plate or where a ladder is available: in these cases, if he chooses to get the grapes, this is an indication that he values them, because the beliefs about the relevant issues in the choice are also easy to elicit. Indeed, even if there were a sophisticated procedure for eliciting attitudes using the preferences in the former situation alone, it is more prone to error and more likely to yield counter-intuitive results than the use of preferences in the latter situation. The recourse to other appropriate situations, permitted by the stability principle, is a more robust way of eliciting preferences.

These considerations yield an elicitation method in which the stability principle plays a central role. We will say that a decision situation is *simple* if the classic elicitation methods apply uncontroversially in this situation; in particular, one expects the decision-maker to maximise expected utility and his utilities to be independent of the states. To elicit the probability and utility of the agent in any given decision situation σ, find another pair of decision situations σ_1 and σ_2 such that (1) σ_1 has the same set of states as σ and σ_2 has the same set of consequences as σ; (2) σ_1 and σ_2 are simple; (3) σ_1 and σ_2 are suitably related to σ in the sense of the stability principle, so that the principle can be taken to apply. To elicit the agent's probabilities in σ, elicit his probabilities in σ_1 using traditional methods (these are robust since σ_1 is simple). By stability, the probabilities elicited in σ_1 are also the agent's probabilities in σ. The agent's utilities in σ can be elicited in a similar way, by eliciting the utilities in σ_2.

This elicitation method, based as it is on stability, does not suffer from the objection to the previous one. If the fox, as he walks away from the grapes, spots the ladder, grabs it and clambers up to get the grapes, this choice is a factor in the elicitation of the attitudes he had as he was walking away. Since there is no reason for him to change his utilities when he spots the ladder, by stability, it can be assumed that they are the same when he walks away (the second situation in the story) and when he returns with the ladder (the third situation). Since the latter situation shares the same set of consequences (obtaining the grapes or not) and is simple (the utilities of the grapes are easily read off from his choices), the utilities elicited in this situation provide a reliable indication of his utilities as he was walking away. According to this process of elicitation, his utilities *are* the same in throughout the story, as the intuition concerning short-term reversibility would suggest. These considerations, taken in tandem with the arguments above in favour of the stability principle, suggest that this is a more adequate method of elicitation in cases where change of attitudes is of interest.

As noted above, a theory of decision should vaunt a representation theorem, which, in many cases, is the theoretical version of the intuitions at work in the elicitation method. Here is no exception: a representation theorem based on exactly this elicitation technique is presented and defended in Hill (forthcoming). Detailed presentation of this result goes beyond the scope of the paper; let us nevertheless draw attention to several aspects which are relevant here. First of all, as anticipated above, the main principles of Savage's theorem are retained: it is assumed that order, independence and continuity hold in all decision situations. By contrast, state-independence is not assumed to hold in all situations; indeed, in the theorem, the simple situations are formally defined as those where the state-independence axioms do hold (and hence, where traditional results apply). The result thus goes beyond Savage in that it can deal with cases where state-independence fails to hold, and it does so by referring to situations where it does hold. As noted above, there are independent arguments against the general validity of state-independence.

The axioms which replace it are "diachronic", and concern the set of decision situations which are suitably related to a given decision situation, in the sense of the stability principle. The first is a richness axiom, requiring that, for any given decision situation, there exists the sorts of situations required to elicit probabilities (resp. utilities), i.e. suitably related simple situations having the same set of states (resp. consequences). The second is a consistency axiom, requiring that if there are several suitably related simple situations which could be used to elicit the probabilities (resp. utilities) in a given decision situation, they give the same result. The consistency axiom is a direct consequence of the stability principle (and the properties of the relation "begin suitably related"): by stability, if simple situations σ_1 and σ_2 are suitably related, the agent has the same probabilities in them; since the situations are simple, these probabilities can be elicited using traditional methods; hence, the probabilities elicited using traditional methods in these two situations must be the same. To this extent, this axiom can be seen as translating into behavioural terms some of the content of the stability principle. The richness axiom can be thought of as a weakening of state-independence: whereas it is largely accepted that the state-independence axioms do not hold in every decision situation, it is equally evident that there are situations in which they do hold, and it is a version of this weaker fact which is required by the richness axiom. For further discussion of these axioms, and their precise formulation in terms of preference orders, see Hill (forthcoming).

Consequences: the situation-dependent factor The stability principle underlies a technique for eliciting the agent's probabilities and utilities which, it has been argued, yields more adequate results than traditional techniques, especially when change of attitudes is at issue. One final question remains to be addressed: what sort of representation of preferences does this elicitation technique provide?

It should be evident that it is not *necessarily* a representation by the traditional expected utility formula (in the Savage framework): $\sum_{s \in S} p(s).u(f(s))$. Consider a non-simple situation σ and appropriately related simple situations σ_1 and σ_2 having the same states and consequences, respectively, as σ. The elicitation method sketched above yields probabilities p and u obtained by applying traditional methods in the latter two situations. There is however no guarantee that p and u represent the preferences in the σ: it could be for example that f is preferred to g in σ, but that $\sum_{s \in S} p(s).u(f(s)) < \sum_{s \in S} p(s).u(g(s))$. For example, if the fox takes the opportunity to grab the grapes when he spots the ladder, one can infer that his utility for the grapes were as high when he was walking away as it was in the first situation when he attempted to get the grapes. If, furthermore, he were offered a bet on the position of the grapes and accepts the same bet before and after his failed attempt, it can be inferred that his beliefs about the states of the world (position of the grapes) have not changed. This would indicate that the utility and the probability in the first situation (where he attempts to get the grapes) and in the second situation (where he walks away) are the same. However, the preferences in these two situations differ, and so cannot both be represented by the expected utility formula involving the elicited probabilities and utilities.

The most natural appropriate form of representation has already been mentioned in this paper: it is the form proposed in Section 2.2.3 which involved the situation- or context-dependent factor γ.[21] Since the Savage principles of order, independence and continuity are satisfied for any situation σ, for any such situation there is a function on state-consequence pairs, call it V, which represents the preferences in σ (Section 2.3.1). As noted above, V may not be equal to the product of the probability and utility elicited by recourse to appropriate simple situations; however, it is always possible, by taking the quotient, to define a factor γ such that $V(s,c) = p(s).u(c).\gamma(s,c)$ for all states s and consequences c.[22] But this just yields the representation proposed in Section 2.2.3: f is preferred to g in σ if and only if $\sum_{s \in S} p(s)u(f(s))\gamma(s, f(s)) > \sum_{s \in S} p(s)u(g(s))\gamma(s, g(s))$.[23] The situation-dependent factor, proposed above as a possible analysis of the phenomenon of sour grapes, comes out as a natural consequence of the elicitation technique.

We thus have a more general motive for introducing and taking account of something of the order of a situation-dependent factor. If one accepts the stability principle as a norm for rational behaviour and a guiding principle for deciding what an agent's beliefs and utilities are, the situation-dependent factor is needed to "fill the gap" between the beliefs and utilities elicited and the agent's preferences in a given

---

[21] Several other ways of representing changes in situation-dependent aspects were discussed in Remark 2, Section 2.2.3. Discussion of the relationship to the change in the set of states goes beyond the current paper; the Jeffrey framework is discussed in Remark 3 below.

[22] For details, see Hill (forthcoming). Naturally, a rather innocent axiom regarding null events is required to ensure that the quotient is well-defined when it should be.

[23] This multiplicative representation is the most natural, given that V is normally decomposed into the product of probabilities and utilities. Other possible representations can be imagined (e.g. δ such that $V(s,c) = p(s).u(c) + \delta(s,c)$), but it is unclear how they would be understood.

situation. In hindsight, this should not be surprising. If the beliefs about the relevant states and the utilities for consequences have not changed, then the only other option is some aspect of the way the agent represents the decision problem to himself. Similarly, if you elicit the agent's probabilities over states and utilities for consequences using a method which does not rely uniquely on the situation under consideration, and if his preferences are not represented by the expected utility calculated with these probabilities and utilities, then you have discovered that there is an aspect of the situation as the agent represents it which still needs to be taken into account. The situation-dependent factor is just a simple way of representing this factor.

Remark 3 To those with sympathies for Jeffrey's decision theory, the conclusions above may be particularly welcome. As we saw in Section 2.2.3, there are analogies between the situation-dependent factor γ and the probability of the consequence conditional on the act and the state. To the extent that conditional probabilities are, for many, a central difference between Jeffrey's decision theory and Savage's (but see Levi 2000; Spohn 1977), the above argument could be considered as a vindication of the former over the latter.

Some important differences between the Jeffrey and Savage frameworks should however be noted. The main problem of this section – the problem of elicitation – receives a better treatment in the latter than in the former, in at least two respects. First of all, the classic representation theorem in the Jeffrey framework (Bolker 1967) does not yield unique probabilities, nor utilities which are as unique as Savage's. This has been corrected, at the expense of extra structure, in the work of Joyce (1999) and Bradley (2007b). Secondly, the representation theorem and framework do not easily lend themselves to practical elicitation of attitudes: not only have there been no attempts to elicit attitudes using the Jeffrey framework, but in work on the subject, there is little indication as to how this could be done in general (for example, in Jeffrey [1972], there are only indications as to how to elicit probabilities on the set of indifferent propositions, but not on propositions in general [Chapter 7]).

Indeed, insofar as the elicitation method proposed above purports to be a method for measuring utilities (of consequences), probabilities (of states) and the situation-dependent factor, it could be seen as a contribution to alleviating this weakness of the Jeffrey framework. Using simple situations to elicit probabilities and utilities does not depend on the use of the Savage framework rather than the Jeffrey framework. The only element of the representation theorem described above which is specific to the Savage framework is the use of Savage's independence axiom (the so-called sure-thing principle) to get a representation by a function V of state-consequence pairs. However, Bolker's axioms provide a function of this sort for the Jeffrey framework (indeed, they provide more than this). It is thus natural to conjecture that a representation theorem of the sort described above, which elicits unique probabilities, unique utilities and probabilities of consequences given acts and states rather than a γ-factor, can be proved in the Jeffrey framework. Further discussion of the interpretation of such a theorem and of these conditional probabilities is beyond the scope of this paper; we will content ourselves with signalling this as an area for future research.

In conclusion, three things can be at the root of a given change in preferences: a change in utilities, a change in beliefs with respect to the states, or a change in some factor in the decision problem. A particularity of the phenomenon of sour grapes is that it is not clear which of these changes is involved. The properties of sour grapes-style changes have not proved sufficient to clarify this issue. To determine which change is involved in a particular case of preference change, one will have to elicit the agent's beliefs and utilities. And in considering what is the best way of doing so, given that the elicitation is to be used to understand change, the situation-dependent factor turns out once again to play a crucial role. So it cannot be dismissed as an easy solution to the problem of preference change, or as a special trick which is only relevant for the case of sour grapes. It poses a challenge to those seeking to understand preference change: to correctly model the agent's representation of the decision problem, its relationship with other attitudes and the changes it undergoes. It is unclear how a model of decision and attitude change which cannot account for this would ever be complete.

Acknowledgements The author would like to thank Philippe Mongin and Francesca Poggiolesi for inspiring discussion and helpful suggestions, and three anonymous referees for their detailed comments.

References

Arrow, K. J. (1974). Optimal insurance and generalized deductibles. *Scandinavian Actuarial Journal*, 1:1–42.

Baltag, A. and Smets, S. (2006). Dynamic belief revision over multi-agent plausibility models. In Bonanno, G., van der Hoek, W., and Wooldridge, M., editors, *Proceedings of the 7th Conference on Logic and the Foundations of Game and Decision Theory (LOFT06)*, pp. 11–24. Amsterdam University Press, Amsterdam.

Bolker, E. (1967). A simultaneous axiomatisation of utility and subjective probability. *Philosophy of Science*, 34:333–340.

Bradley, R. (2005). Radical probabilism and mental kinematics. *Philosophy of Science*, 72:342–364.

Bradley, R. (2007a). The kinematics of belief and desire. *Synthese*, 56:513–535.

Bradley, R. (2007b). A unified bayesian decision theory. *Theory and Decision*, 63:233–263.

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

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

Gärdenfors, P. (1988). *Knowledge in Flux : Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA.

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

Hill, B. (2007b). The logic of awareness change. In Arrazola, X. and Larrazabal, J. M., editors, *Proceedings of the First ILCLI International Worskop on Logic and Philosophy of Knowledge, Communication and Action*. University of the Basque Country Press.

Hill, B. (2008a). Towards a "sophisticated" model of belief dynamics. Part I: The general framework. *Studia Logica*, 89(1):81–109.

Hill, B. (2008b). Towards a "sophisticated" model of belief dynamics. Part II: belief revision. *Studia Logica*, 89(3):291–323.

Hill, B. (forthcoming). Living without state-independence of utilities. *Theory and Decision*.

Jeffrey, R. C. (1972). *The Logic of Decision*, 2nd edn. University of Chicago Press, Chicago.

Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge.

Kahneman, D., Wakker, P. P., and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics*, 112:375–405.

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

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

Levi, I. (2000). Review of *the foundations of causal decision theory*. *The Journal of Philosophy*, 97:387–402.

Modica, S. (2008). Unawareness, priors and posteriors. *Decisions in Economics and Finance*, 31.

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

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

Spohn, W. (1977). Where luce and krantz do really generalize savage's decision model. *Erkenntnis*, 11:113–134.

van Benthem, J. and Liu, F. (2007). Dynamic logic of preference upgrade. *Journal of Applied Non-classical Logic*, 17:157–182.

van Ditmarsch, H., van der Hoek, W., and Kooi, B. (2007). *Dynamic Epistemic Logic*. Springer, Dordrecht.