Utility Independence in a Qualitative Decision Theory

Fahiem Bacchus
Department of Computer Science
University of Waterloo
Waterloo, Ontario
Canada, N2L 3G1
fbacchus@logos.uwaterloo.ca

Adam J. Grove
NEC Research Institute
4 Independence Way
Princeton NJ 08540, USA
grove@research.nj.nec.com

Abstract

Qualitative accounts of utility modeling and decision theory offer the prospect of reasoning about preference and decision-making without requiring hard-to-obtain numerical probabilities and utilities. It is plausible that such accounts can be found because qualitative criteria (in particular, *dominance*) seems to play a large role in human decision making; the formal quantitative apparatus of *maximum expected utility* tends to be invoked only in the most critical, most finely-balanced, cases.

In this paper, we show how non-probabilistic *independence* concepts—such as preferential independence and utility independence—can be integrated with other sources of quantitative information. It turns out that there are some subtleties involved in making sense of these ideas in a logical framework. The main contribution of this paper is to demonstrate these subtleties, and then give semantics that avoid many of the problems. We then argue that knowledge of utility independence can be a useful addition to the qualitative reasoner's tool-kit.

1 Introduction

It is often suggested that rational decision-making should be based on the principle of *maximum expected utility* (MEU). To use MEU one must have a numeric *utility* function that quantifies how desirable or undesirable each particular state of the world is, and a family of probability distributions over the states of the world. This family of distributions is indexed by the actions one might take, each distribution telling us the probability that any particular state of the world will be brought about by the associated action. The MEU principle advocates selecting that action which leads to the greatest expected utility. For an introduction see, for example, [Fre88, GS88, Sav54].

However, it is well known that there are several epistemological and computational difficulties involved in using MEU. In particular, it is often extremely difficult or even impossible to obtain the probability and utility functions required. People tend to express their beliefs, goals, and preferences in different, generally qualitative, terms and have trouble translating these into numerical distributions and utility functions. Even when it is possible to obtain them, it might not be practical to use numeric probabilities and utilities directly. For instance, if there are n independent Boolean propositions the state space may have size 2^n, so that an explicit listing of probabilities or utilities quickly becomes unmanageable. Furthermore, such a listing might obscure valuable structure or heuristic information that is apparent in a more "natural" specification. For example, we might lose the ability to quickly recognize when dominance arguments render detailed utility calculations redundant.

Such problems are one motivation for recent interest in *qualitative* theories of probability and uncertainty, utility (i.e., preference), and their combination (i.e., decision making). There has been greatest progress on qualitative theories of probability. This includes theories of probabilistic independence (notably Bayes nets [Pea88]) which use qualitative information to simplify the acquisition and use of numerical probabilities, theories of extreme probabilities such as ϵ-semantics [Pea89], and the related area of non-monotonic reasoning [Gin87] which (according to some interpretations) seeks to replace probabilities entirely with a qualitative counterpart.

Work concerning qualitative theories of utilities and decision theory is more recent, and far less developed.[1] Among the papers of direct relevance to us include

[1] However, we acknowledge the large body of existing work on *deontic logics* (e.g., [von51]) and *preference logics* (e.g., [von72]). This appears to be only weakly relevant to our work, however, because these logics do not generally appeal to the expected-utility paradigm.

those by Boutilier [Bou94], Pearl and Tan [Pea93, TP94a, TP94b], and Doyle, Wellman, and Shoham [DW91, DW94, DSW91]. We discuss these and other papers in more detail in Section 3.

In some previous works the approach taken has been to dispense with numeric utility and probability functions, instead replacing them with qualitative analogs. For instance, [Pea93] suggests using probabilities of the form ϵ^k for natural numbers k and ϵ a small positive number, and utilities of the form $\pm(1/\epsilon)^k$. By considering the limit $\epsilon \to 0$ (i.e., where probabilities are "very small", utilities are "very large", and all we care about are order-of-magnitude distinctions) one can hope to simplify the reasoning process. There are many interesting variants of this basic idea, including the use of qualitative probabilities alone (such as κ-rankings [Pea93]) or qualitatively ranked utilities alone [TP94a].

In this paper we also consider the problem of decision-making using qualitative, or limited amounts of quantitative, information. An important contrast with previous work is that we will not assume that probabilities or utilities are themselves qualitative (although they may be). Instead, our goal is to work towards a decision theory that can handle qualitative *knowledge* about probabilities and utilities. The particular focus of this paper is knowledge about *independence* for utilities and preferences. Whether or not the probabilities and utilities about which we make independence assertions are in any sense qualitative is generally an orthogonal issue.[2] In Section 4 we give a result illustrating one way in which these two notions can be usefully combined.

We have investigated independence concepts for preference and utility in recent work [BG95]. Most of these concepts have been known for a long time, albeit perhaps not in the A.I. literature, as part of an area known as *multi-attribute utility theory* [KR76]. In [BG95] we were interested in the possibility of graphical models for these concepts, analogous to graphical techniques for probabilistic independence (such as Bayes Nets). In this paper we suggest a different (although related) use for these concepts: that independence assertions are an important source of information about utilities.[3] Like probabilistic independence, these notions are qualitative, relatively easy for people to access, and can simplify computation. Thus, they can help in achieving a more usable version of (qualitative) decision theory.

There are perhaps two reasons why independence concepts are relatively unexplored in A.I. One, which we will address later in the paper, is simply that they are relatively weak (in contrast to the theory of probabilistic independence, which is mathematically richer). Another is that they have only been defined in a rather simple context, involving product spaces of attributes. In contrast, much of the work in qualitative decision theory uses concepts from logic. Standard multi-attribute utility theory might consider a space described by several attributes including, for example, *health* and *wealth*. The standard theory can make sense of the assertion that, for instance, one's *health* is utility independent of the set of all other attributes (including *wealth*). But the standard formulation would have problems saying 1) one's *health* is utility independent of *wealth simpliciter*, or 2) that the logical sentence *health* $\vee$ *wealth* is independent of everything else, or 3) coping with logical constraints, such that the lowest level of *wealth* is incompatible with the highest level of *health*.

A principal contribution of this paper is to show how to define the standard independence notions in a logically rich context. Although our approach is not technically complex, it has several interesting (and possibly controversial) philosophical aspects, because there are several other definitions one might use. The heart of this paper is Section 3, which presents our proposal and discusses some of the difficulties it tries to address. This section also contains some more detailed comparisons with related work. Section 4 explores further consequences of our definitions. In particular, we state results showing how certain independencies interact with other pieces of qualitative information. For example, there are many cases in which the dominance arguments that one might wish to use are invalid unless appropriate independencies are given. In Section 2 we present the basic background material.

2 Standard Independence Concepts

As we have said, we assume familiarity with the basic ideas and techniques of decision theory and the expected-utility paradigm. In particular, it is outside the scope of this paper to defend the MEU principle or to examine the many alternative decision theories that are occasionally proposed. The purpose of this section is to establish some notation and to then give a brief survey of independence concepts for preference and utility.

We often assume that the set of states S is defined by a set of binary attributes (i.e., propositional variables) $V = \{p_1, \ldots, p_n\}$. Hence, S can be considered to be the set of all truth assignments over V. Furthermore, we can use propositional logic to talk about events over S. In general, it is very useful to also allow non-binary attributes, corresponding to concepts or resources that have more than two levels. Of course, it is trivial to

[2] Note that the phrase "qualitative decision theory" is occasionally used to denote theories in which probabilities and/or utilities are themselves qualitative. This is not our usage here.

[3] Unlike probabilistic independence, there are a number of distinct notions of independence relevant to preference and utility. We will simply use the term "independence" to refer to any of these notions when the distinction is not important.

formulate a "propositional"-style logic which can talk about such attributes as well. Everything in this paper applies whenever all attributes take on a discrete number of values; we will to present our results and discussion in terms of binary-valued attributes, but this is solely for notational simplicity. Continuous or real-valued attributes raise distinct issues, and so will not be considered in this paper.

In the following, if $X \subseteq V$ then $f(X)$ stands for some real valued function all of whose arguments are in X, i.e., $f(X) : 2^X \longrightarrow I\!R$ is a function that depends on the truth value of the variables in X only. A *utility function*, u, is a function over complete states, i.e., $u(V)$, and thus it can potentially take on exponentially many unrelated values, one for every state.

We will also need to refer to probability distributions over S. More generally, when $X \subseteq V$ and we say that Pr is a probability distribution over X, we mean that Pr is a probability distribution over the set of truth assignments to the variables in X.

A utility function u induces a *preference ordering* $\succeq$ on the probability distributions over S as follows:

$$Pr_1 \succeq Pr_2 \quad \textit{iff} \quad \sum_{s \in S} Pr_1(s)u(s) \geq \sum_{s \in S} Pr_2(s)u(s),$$

where Pr_1 and Pr_2 are two distributions over S. That is, we prefer Pr_1 to Pr_2 if Pr_1 induces greater expected utility. Thus utility serves to characterize not only the agent's values but also its attitudes towards risk: it ranks probabilistic gambles between various outcomes.

Sometimes, instead of considering a preference ordering over probability distributions on states, we are only interested in the order among the states themselves. In particular, note that any utility function u induces a unique preference ordering over the individual states: $s \succeq s'$ for two states s and s' iff $u(s) \leq u(s')$. But the converse is not true: since utility functions also reflect one's attitude towards risk, many distinct utility functions can lead to the same preference ordering over states.[4]

We now briefly summarize a number of standard independence notions for utility and preference. With the exception of the final definition (conditional additive independence [BG95]) these are standard ideas from the field of *multi-attribute decision theory*. This review is based on the following sources [Fis82, Fre88, KR76, KLST71] and our paper [BG95].

The first definition we give is that of *preferential independence*. This is the weakest notion we discuss because it only considers the preference ordering among individual states. Intuitively, a set of attributes X is preferentially independent of everything else, if when we hold everything else fixed (i.e., the values of attributes $V-X$), the induced preference ordering over assignments to X does not depend on the particular values that $V-X$ are fixed to. Thus, we can assert that preferences over X hold *ceteris paribus*—i.e., all else being equal.

Definition 2.1 : Recall that each state $s \in S$ is a truth assignment to the variables in V. If $X \subset V$ we can write s as (α, γ), where α is a truth assignment to the variables in X and γ is a truth assignment to the remaining variables $V - X$.

The set of attributes X is *preferentially independent* of $V-X$ when one's preference order among truth assignments to X does not depend on the particular values that the variables $V-X$ are set to. That is,

$$\forall \gamma, \gamma' \in \textit{"truth assignments over } V - X\textit{"} : \\ (\alpha, \gamma) \succeq (\beta, \gamma) \text{ iff } (\alpha, \gamma') \succeq (\beta, \gamma'),$$

where α and β are any two truth assignments to the variables in X. ∎

The concept of *utility independence* is similar, but somewhat stronger because it is concerned with the induced utility function (and not just preferences between individual states). Thus, the relative strength of preference between states (and not just the order of these preferences) must stay the same. Put another way, one's attitude towards risk should not change. Like preferential independence, utility independence can also be viewed as a formalization of *ceteris paribus*.

Definition 2.2: Let $X \subset V$ be some set of attributes, and suppose γ is a truth assignment to the remaining variables $V - X$. Given a probability distribution Pr over X, there is a unique distribution Pr^γ over V such that (1) Pr^γ's marginal over X is Pr, and (2) Pr^γ gives probability 1 to γ.

Given a utility function with associated preference ordering $\succeq$, we define the *conditional preference over X given γ*, $\succeq_\gamma$, to be the preference ordering such that

$$Pr_1 \succeq_\gamma Pr_2 \quad \text{iff} \quad Pr_1^\gamma \succeq Pr_2^\gamma,$$

where Pr_1 and Pr_2 are any two distributions over X. ∎

Definition 2.3 : The set of attributes X is *utility independent* of $V-X$ when conditional preferences over X do not depend on the particular value given to $V-X$. That is,

$$\forall \gamma, \gamma' \in \textit{"truth assignments over } V - X\textit{"} : \\ Pr_1 \succeq_\gamma Pr_2 \text{ iff } Pr_1 \succeq_{\gamma'} Pr_2,$$

where Pr_1 and Pr_2 are any two distributions over X. Here $\succeq_\gamma$ and $\succeq_{\gamma'}$ are the conditional preferences over X given γ and γ' respectively. ∎

[4] If u' is a monotonic function of u, then u and u' will lead to the same preference ordering over states. But only if u and u' are linearly related are they equivalent as *utility* functions (i.e., generate the same preferences over probability distributions).

It is worth noting that, if we are just concerned with a single binary attribute being independent of all the others, then utility independence and preferential independence coincide, but this is not the case when the set X contains more than one attribute or if the attributes can take more than one value. In the single-attribute binary case (only), both preferential and utility independence reduce to the particular formalization of *ceteris paribus* given in [DW91].

In general, preferential (resp., utility) independence fails to hold if one has a preference reversal among values of (resp., probabilistic mixtures of values of) the attributes X, when some set of attributes in $V-X$ is changed. Judgments of utility independence and preferential independence appear to be fairly natural and common; see [KR76] for a very extensive discussion. They are, at heart, judgments about *relevance* and people seem to be fairly good at this in general.

We close by considering an even stronger notion: *additive independence*.

Definition 2.4: Let $Z_1, \ldots, Z_k$ be a partition of V. $Z_1, \ldots, Z_k$ are additively independent (for $\succeq$) if, for any probability distributions Pr_1 and Pr_2 that have the same marginals on Z_i for all i, Pr_1 and Pr_2 are indifferent under $\succeq$, i.e., $Pr_1 \succeq Pr_2$ and $Pr_2 \succeq Pr_1$. ∎

In other words, one's preference only depends on the marginal probabilities of the given sets of variables, and not on any correlation between them. Conditional versions of both additive and utility independence can be defined. The definitions require that the specified independence holds whenever some subset of variables are held fixed. For instance, the following independence concept was developed in our earlier work [BG95].

Definition 2.5: X and Y are *conditionally additively independent* given Z (X, Y, Z disjoint, $X \cup Y \cup Z = V$) iff, for any fixed value γ of Z, X and Y are additively independent in the conditional preference structure over $X \cup Y$ given γ. ∎

All of these notions of independence have interesting consequences for the form of the utility function. Although knowing these consequences might help somewhat in understanding our results in Section 4, they are not essential and so we omit them. Just to give a flavor, though, here is one of the strongest and most important:

Proposition 2.6: *([KR76]) $Z_1, \ldots, Z_k$ are additively independent for $\succeq$ iff the utility function representing $\succeq$ can be written as*

$$u(V) = \sum_{i=1}^{k} f(Z_i)$$

for some functions f_i.

3 Independence for Formulas

Most existing research in qualitative decision theory is concerned with assertions about logical formulas. For instance, both [Bou94] and [TP94a] give semantics to the assertions of the form "if ψ is known then φ is preferred to $\neg\varphi$", where φ and ψ can propositional logic formulas. The related area of deontic logic also supposes that one should reason about preference and obligation in a logical setting.

In contrast, the various definitions of independence given in Section 2 only deal with attributes (which for us tend to be individual propositional variables) or sets of attributes, not arbitrary formulas. This is a significant restriction, and the purpose of this section is to show how it can be relaxed.

To see part of the difficulty caused by formulas, first consider the simple case where one variable (p_1 say) is preferentially independent of the remaining variables. To simplify the discussion, suppose that the direction of the preference is towards p_1 being true. Then the definition of preferential independence says that, for every pair of states s, s' that agree on the values given to all the $n-1$ remaining variables, we will always prefer the one in which p_1 is true.

Why is this case so straightforward? There are two distinct reasons. First, it seems to be a reasonable formalization of the idea that p_1 is preferred to $\neg p_1$ *ceteris paribus*, i.e., preferred given that "all else is equal". The point is that there is little doubt as to what "all else" should refer to: we should fix the values of all the propositional variables other than p_1. Second, once we fix the values of the other variables, we are left with only two states: one satisfying p_1 and the other $\neg p_1$. There is no doubt as to what "preference" means here: the former state should have higher utility than the latter.

But now suppose that, instead of a primitive propositional variable, it is an arbitrary logical formula that is "preferred" to its negation. To be concrete, we consider the formula $\varphi = p_1 \otimes p_2$ (i.e., the exclusive-or of p_1 and p_2.) What is the "all else" that we are supposed to hold fixed when comparing φ with $\neg\varphi$? There is no clear answer to this. Furthermore, as we see shortly, once we have fixed "all else", we may be left with more than one φ state and more than one $\neg\varphi$ state. How should we compare them?

Such questions have been considered by Doyle, Shoham, and Wellman [DSW91], and also by Tan and Pearl [TP94b]. Roughly speaking, in the case of $\varphi = p_1 \otimes p_2$ they would fix the values of all propositional variables other than p_1 and p_2. (In general, they fix all propositional variables that are not required to appear in the formula being considered. This is how the *ceteris paribus* condition is interpreted.) Note that, for any fixed values of the other variables, we are left

with a set of four states, corresponding to the four truth assignments to p_1 and p_2. Their interpretation of preference is that, among each such set of four states, the two in which φ is true are preferred to the other two, in which φ is false. (That is, *each* state satisfying φ has higher utility than *both* of the $\neg\varphi$ states.)

In Section 3.1, we argue against this interpretation of preference. Instead, we endorse an interpretation proposed by Jeffrey [Jef65] that is based on the idea of *desirability* or (as we prefer to call it) *conditional expected utility*. We also disagree with the [DSW91, TP94b] interpretation(s) of *ceteris paribus*. In Section 3.2 we present our concerns and give an alternative approach. Our approach uses conditional expected utility as the base semantics for preference and to make sense of the ideas of utility (resp., preferential, additive) independence over arbitrary collections of formulas.

3.1 Conditional Expected Utility

Given that one has a collection of φ states and $\neg\varphi$ states, what does it mean that the former are preferred to the latter? The [DSW91, TP94b] proposal is that all of the former are preferred to all of the latter. This is an extremely strong condition with undesirable consequences.

One important problem is that it becomes impossible to override preferences given more specific information.[5] One cannot say, for instance, that φ is preferred to $\neg\varphi$ and at the same time that, conditioned on some other information ψ, we prefer $\neg\varphi$ to φ. However, the pattern in which a general preference is overridden by its reverse in more specific situations occurs frequently. For example, there is a preference for not having surgery over having surgery, yet in the circumstance where surgery would improve one's long term health this preference might be reversed.[6] Hence, it is essential to be able allow for preference overriding. We note that Tan and Pearl, in [TP94a], acknowledge this and propose a modification to their earlier theory that allows statements about overriding preferences. However, their proposal essentially amounts to the simple stipulation that one should ignore general preferences when they are overridden: the underlying semantics are not changed. This seems unsatisfactory to us. Furthermore, if the underlying semantics is incompatible with such a basic pattern of preference, then one can have little confidence that this is the only problem.

[5] See [TH96] for other criticisms of these semantics for preference.

[6] Note that stating that this preference holds *ceteris paribus* does not address the problem. The assertion that the preference holds *ceteris paribus* still means that it is required to hold under any *fixed* setting of the other conditions. So given the fixed condition of needing surgery, these semantics still force a preference for not having surgery over surgery.

Instead, we prefer Jeffrey's proposal from [Jef65], which we refer to as *conditional expected utility*. This is defined if one has a probability function Pr over the underlying space S. Then the conditional expected utility over any subset $T \subseteq S$ can be defined as

$$U(T) = \frac{\sum_{t \in T} Pr(t)\, u(t)}{Pr(T)} \tag{1}$$

where we use U to denote the aggregate utility function. Thus, if the collection of states satisfying φ has higher conditional expected utility than the collection of states satisfying $\neg\varphi$, then we assert that φ is preferred $\neg\varphi$.

In general, if φ and ψ are arbitrary formulas, then we write $\varphi \succeq \psi$ to assert that $U(\varphi) \geq U(\psi)$, where we identify a formula with the set of states satisfying it. (Similarly, $\varphi \succ \psi$ just if $U(\varphi) > U(\psi)$.) Conditional preferences are also easy to interpret: $\varphi_1 \succ \varphi_2$ *given* ψ means that $U(\varphi_1 \wedge \psi) > U(\varphi_2 \wedge \psi)$. It is easy to see this semantics is compatible with statements involving overridden preferences. For instance, the two statements $\varphi \succ \psi$ and $\psi \wedge \omega \succ \varphi \wedge \omega$ can be consistently asserted together.

Perhaps the best intuitive reading of preferences based on conditional expected utility is that they correspond to how one might react to various pieces of news. If $\varphi \succ \psi$ one should be happier to hear that φ is true than to hear that ψ is true. (In Section 5 we briefly discuss how actions might be introduced.) In this paper, we will not give any further defense of Jeffrey's semantics for utility aggregation, mostly because many of the best arguments are in his book [Jef65].

Another important difference between the notion of $\succeq$ just defined and those proposed in [DSW91, TP94b] is that we have not (as yet) invoked any form of *ceteris paribus* condition. In contrast, as discussed earlier, when [DSW91, TP94b] assert $\varphi \succeq \psi$ they are making an assertion that holds (roughly speaking) when other propositions are fixed, i.e., holds *ceteris paribus*. To express *ceteris paribus* conditions in our context, we provide a general mechanism where by a variety of utility independence assertions can be stated. These assertions can be (but need not be) stated independently of assertions about preference. We present the details of this proposal in the next section.

3.2 Independence and Ceteris Paribus

[DSW91, TP94b] give semantics to preference statements that embeds a notion of *ceteris paribus*. In particular, their interpretation of *ceteris paribus* involves considering fixed values for all the propositional variables not mentioned in the formulas being considered.

One problem is that such semantics are very syntax dependent, and thus the conclusions they support can be rather arbitrary. To see why, consider again the

formula $p_1 \otimes p_2$ (exclusive-or) and suppose that we redefine our vocabulary so that p_1 is replaced by a new propositional symbol p_1', such that $p_1' \equiv p_1 \otimes p_2$. In this new language, the old p_1 would be expressed using a compound sentence; in fact, $p_1 \equiv p_1' \otimes p_2$. Since p_1 and p_1' are interdefinable, the new vocabulary is just as expressive as the old, and so it may only be a matter of convention as to which is used. Yet preferential independence of φ is given two different meanings according to whether p_1 or p_1' is primitive.

An even more important problem is that such semantics are inflexible. These semantics commit to a single, fixed, interpretation of *ceteris paribus* that applies to all assertions about preferences. For instance, these semantics do not easily allow one to say that $p_1 \succeq \neg p_1$ independent of the value of p_2 and p_3, while at the same time allowing this preference to possibly be *dependent* on the value of p_4.

Our proposal, which avoids these problems, depends on the concept of the set of *atoms* formed from a collection of formulas, defined as follows.

Definition 3.1 : If Ψ is a set of formulas, the *atoms* of Ψ is the set of all *consistent* conjunctions that can be formed from the members of Ψ by including each $\psi \in \Psi$ or its negation. For example, if $\Psi = \{p, q \wedge r\}$ then the atoms of Ψ are $\{p \wedge (q \wedge r)$, $\neg p \wedge (q \wedge r)$, $p \wedge \neg(q \wedge r)$, $\neg p \wedge \neg(q \wedge r)\}$. ∎

For any k formulas, there will be (at most) 2^k atoms. We say "at most" because all combinations might not be consistent, and in this paper we restrict the term "atom" to logically consistent formulas.

The collection of atoms over any set of formulas can be thought of as a new space of states, in which the given formulas play the role of primitive attributes. Each of these atoms corresponds, in general, to a collection of states from the original state space. From the previous section, we know that it is possible to give any collection of states a "utility" value using the idea of conditional expected utility. Thus, an induced utility function can be defined over the space of atoms. Any assertion of utility independence involving the collection of formulas can now be interpreted as an assertion about this induced utility function. Since the formulas are primitive attributes in the new state space, we can use the standard definitions to interpret these independence assertions.

More formally, let $\Psi = \{\psi_1, \ldots, \psi_k\}$ be a collection of formulas. Let the underlying space be S, with utility function u and probability distribution Pr. Consider the set of atoms of Ψ. Each such atom corresponds to a consistent truth assignment to the formulas in Ψ, where the formula ψ_i is assigned the value true just if it appears positively in the atom. We define a new space S^Ψ consisting of all of these truth assignments. A utility function u^Ψ over S^Ψ is defined using conditional expected utility. Specifically, the utility u^Ψ of a state s in S^Ψ is defined to be the conditional expected utility, in the original space, of the atom that corresponds to s. Similarly, a probability distribution Pr^Ψ over S^Ψ is defined using marginalization. That is, Pr^Ψ of a state s in S^Ψ is the probability under Pr (i.e., in the original space) of the set of worlds satisfying the corresponding atom.

For example, if $\Psi = \{\psi_1, \psi_2, \psi_3\}$ then S^Ψ will be the set of 8 truth assignments to the ψ_i (assuming that all atoms are consistent). Thus, using the above definitions, $u^\Psi(\psi_1 \wedge \neg\psi_2 \wedge \psi_3) = U(\psi_1 \wedge \neg\psi_2 \wedge \psi_3)$. (Recall that U is defined by Equation 1.) Note that here we write the atom itself to refer to the corresponding truth assignment. Similarly, $Pr^\Psi(\psi_1 \wedge \neg\psi_2 \wedge \psi_3) = Pr(\psi_1 \wedge \neg\psi_2 \wedge \psi_3)$.

Using the above correspondences, we interpret assertions of independence among a set of formulas Ψ as making assertions about the utility and probability functions on the induced space S^Ψ. Since, the formulas of Ψ are primitive attributes in the induced space, the standard definitions given in Section 2 can be applied almost without change. The only difference arises because not all possible truth assignments are consistent.

Example 3.2: Suppose that we wish to assert that the set of formulas $\{p, q\}$ is preferentially independent of the formula $p \vee (q \wedge r)$. Then we let $\psi_1 = p$, $\psi_2 = q$, $\psi_3 = p \vee (q \wedge r)$ and $\Psi = \{\psi_1, \psi_2, \psi_3\}$. If we were to ignore the issue of inconsistent atoms and apply Definition 2.1 literally, then this assertion states that the preference ordering among the truth assignments (written as atoms) $\{\neg\psi_1 \wedge \neg\psi_2 \wedge \neg\psi_3$, $\neg\psi_1 \wedge \psi_2 \wedge \neg\psi_3$, $\psi_1 \wedge \neg\psi_2 \wedge \neg\psi_3$, $\psi_1 \wedge \psi_2 \wedge \neg\psi_3\}$ must be the same as that among the truth assignments $\{\neg\psi_1 \wedge \neg\psi_2 \wedge \psi_3$, $\neg\psi_1 \wedge \psi_2 \wedge \psi_3$, $\psi_1 \wedge \neg\psi_2 \wedge \psi_3$, $\psi_1 \wedge \psi_2 \wedge \psi_3\}$. That is, the truth or falsity of ψ_3 should not affect one's preferences between the various valuations of ψ_1 and ψ_2.

However, this is not entirely meaningful: $\psi_1 \wedge \neg\psi_2 \wedge \neg\psi_3$ and $\psi_1 \wedge \psi_2 \wedge \neg\psi_3$ are both inconsistent (since if $\psi_1 = p$ is true then $\psi_3 = p \vee (q \wedge r)$ must be as well), and so are not part of the space S^Ψ. To address this, we weaken the definition of preferential independence slightly to simply require that all induced orderings be consistent with each other. In this example, the preference ordering between $\neg\psi_1 \wedge \neg\psi_2 \wedge \neg\psi_3$ and $\neg\psi_1 \wedge \psi_2 \wedge \neg\psi_3$ must be the same as between $\neg\psi_1 \wedge \neg\psi_2 \wedge \psi_3$ and $\neg\psi_1 \wedge \psi_2 \wedge \psi_3$. However the preferences between the first two atoms and the atoms $\psi_1 \wedge \neg\psi_2 \wedge \psi_3$ and $\psi_1 \wedge \psi_2 \wedge \psi_3$ are not constrained by this assertion. ∎

Definition 3.3 : (Preferential independence for formulas.) Let $\Psi = \psi_1, \ldots, \psi_j, \psi_{j+1}, \ldots \psi_k$. The set of formulas $\psi_1, \ldots, \psi_j$ is *preferentially independent* of $\psi_{j+1}, \ldots, \psi_k$ when one's preference order among the truth assignments to $\psi_1, \ldots, \psi_j$ consistent with particular values given to the remaining formulas, does

not depend on the values given to these other formulas.

Formally: For any α, β that are truth assignments to $\psi_1, \ldots, \psi_j$, and γ, γ' that are truth assignments to $\psi_{j+1}, \ldots, \psi_k$, then, if *all* four combinations $\{(\alpha,\gamma), (\alpha,\gamma'), (\beta,\gamma), (\beta,\gamma')\}$ are logically consistent, we must have:

$$(\alpha,\gamma) \succeq (\beta,\gamma) \textit{ iff } (\alpha,\gamma') \succeq (\beta,\gamma')$$

where $\succeq$ is the preference relation induced by u^Ψ. ∎

The modified definition of utility independence is sufficiently similar in spirit that we do not repeat it here. The definition of additive independence, Definition 2.4, can by applied without any change in wording. Many of the interesting properties of these independence concepts can be shown to carry over to the new definitions. For instance, we note that the analog of Proposition 2.6 still holds.

With these definitions we have the flexibility to make independence assertions entirely separately from statements about the direction of preference. But, as [DSW91, TP94b] have recognized, it is often convenient to be able to assert both together. Suppose we wish to assert, for instance, that φ is utility independent of ψ_1, ψ_2 and that (no matter what particular values we give to ψ_1 and ψ_2) we prefer φ to $\neg\varphi$. To do this, one could assert the utility independence and then state the direction of preference relative to any single (arbitrary) consistent valuation for $\{\psi_1, \psi_2\}$. For instance:

$$\varphi \wedge \psi_1 \wedge \psi_2 \succeq \neg\varphi \wedge \psi_1 \wedge \psi_2,$$

together with the assumption of utility independence, implies that, e.g.,

$$\varphi \wedge \neg\psi_1 \wedge \neg\psi_2 \succeq \neg\varphi \wedge \neg\psi_1 \wedge \neg\psi_2$$

(and similarly for any other consistent valuation for ψ_1 and ψ_2). But it is useful to create a more natural notation for such cases, which avoids the need to choose an arbitrary valuation for ψ_1 and ψ_2. We interpret an expression of the form

$$\varphi_1 \succeq_{\psi_1,\ldots,\psi_k} \varphi_2$$

as asserting (1) that $\{\varphi_1, \varphi_2\}$ is independent of $\{\psi_1, \ldots, \psi_k\}$, and (2) that, conditioned on any fixed consistent valuation of $\{\psi_1, \ldots, \psi_k\}$, φ_1 has higher conditional expected utility than φ_2.

It should be noted that $\varphi_1 \succ_{\psi_1,\ldots,\psi_k} \varphi_2$, does not entail $\varphi_1 \succ \varphi_2$, nor does the converse hold (even in the presence of utility independence). That is, it is possible to partition the state space and assert that $\varphi_1 \succ \varphi_2$ in every member of the partition, yet simultaneously assert that $\varphi_2 \succ \varphi_1$ unconditionally. Consider, for example, the case where $\varphi \succeq \neg\varphi$ *given* ψ and $\varphi \succeq \neg\varphi$ *given* $\neg\psi$. To see why $\varphi \succ \neg\varphi$ need not hold unconditionally, suppose that ψ is a very much more desired alternative to $\neg\psi$ than φ is to $\neg\varphi$, and that ψ and $\neg\varphi$ are strongly correlated, so that when $\neg\varphi$ is true ψ tends to be true also. Then we would much prefer to learn $\neg\varphi$ than φ if this is *all* we learn (hoping of course that ψ is also true). But, if we know the value of ψ (no matter whether we know it to be true of false), we would prefer φ.

If p_1 is a basic proposition, the [DSW91] interpretation of p_1 being preferred to $\neg p_1$ can be written as

$$p_1 \succeq_{p_2,\ldots,p_n} \neg p_1$$

using our notation (where $p_2, \ldots, p_n$ are the rest of the basic propositions). But our proposal is far more general than this, because there is freedom to use other collections of formulas instead of $\{p_2, \ldots, p_n\}$.[7] Indeed, we can assert several comparisons between p_1 and p_2 simultaneously, each relative to a different set of formulas. Of course, as we have noted, we also have the ability to state independence (of various types) independently of any specific preferential comparison. Finally, note that our proposal has no built-in syntax dependence. One can, and must, explicitly decide what formulas are actually relevant to a comparison.

The key to understanding how one can reason with a collection of independence assertions is to realize that assertions of independence involving formulas impose algebraic constraints on both the utilities and the probabilities over the original space.

Example 3.4: Let the basic propositions be p, q, and r. The original space then consists of 8 states, and can be specified by 8 basic probabilities p_{pqr}, $p_{\bar{p}qr}$, $\ldots$, $p_{\bar{p}\bar{q}\bar{r}}$ and the 8 basic utilities u_{pqr}, $u_{\bar{p}qr}$, $\ldots$, $u_{\bar{p}\bar{q}\bar{r}}$.

Consider the assertion that p is utility independent of $q \wedge r$. According to our semantics, this means that we consider the four atoms of the set $\{p, q \wedge r\}$ (Defn. 3.1). Each atom is attributed a utility as determined by Equation 1. The definition of utility independence reduces in this case to the assertion that that $U(p \wedge (q \wedge r)) - U(\neg p \wedge (q \wedge r))$ have the same sign as $U(p \wedge \neg(q \wedge r)) - U(\neg p \wedge \neg(q \wedge r))$. This is equivalent to the assertion that

$$u_{pqr} - u_{\bar{p}qr}$$

and

$$\frac{p_{p\bar{q}r}u_{p\bar{q}r} + p_{pq\bar{r}}u_{pq\bar{r}} + p_{p\bar{q}\bar{r}}u_{p\bar{q}\bar{r}}}{p_{p\bar{q}r} + p_{pq\bar{r}} + p_{p\bar{q}\bar{r}}} - \frac{p_{\bar{p}\bar{q}r}u_{\bar{p}\bar{q}r} + p_{\bar{p}q\bar{r}}u_{\bar{p}q\bar{r}} + p_{\bar{p}\bar{q}\bar{r}}u_{\bar{p}\bar{q}\bar{r}}}{p_{\bar{p}\bar{q}r} + p_{\bar{p}q\bar{r}} + p_{\bar{p}\bar{q}\bar{r}}}$$

have the same sign. That is, it reduces to an algebraic constraint over the utilities and probabilities of the original space. ∎

[7] We note that [DW94] have a proposal that allows some more flexibility than [DSW91], but it still only allows one interpretation of *ceteris paribus* to apply to any particular set of formulas. Furthermore, their interpretation is built into the semantics of preference assertions, and cannot be modified by assertions in the language they present.

The fact that an assertion about utilities also constrains probabilities may seem surprising, but makes sense philosophically. As we have said, the basic independence concept is *ceteris paribus*. But the condition that "everything else be the same" except for the formula of interest (φ say) is unrealistic. It makes more sense to think of everything else being *as similar as possible* given that φ changes truth value. This phrasing makes the similarity to counterfactual and conditional logic clear (see for instance [Lew73]). In counterfactual logic, for instance, one is interested in what would happen if some assertion were to be true even though it is known to be false. There is general agreement that the appropriate semantics for counterfactuals and conditionals should not consider all the states in which φ is true, but only the most "normal" such states. So we should not be surprised if a robust formalization of *ceteris paribus* should also need a notion of how plausible particular states are. And this is precisely the role of probabilities—to tell us how likely or unlikely we consider various states to be.[8]

Standard independence definitions do not *appear* to be invoking anything other than utilities or preference. However, this is somewhat misleading because information about the similarity of states is hidden in the choice of attributes or *framing* [DW91]. [DW94] discuss this further, and also argue that making sense of *ceteris paribus* requires more structure than just the utilities (unlike us, however, they do not suggest probabilistic semantics). [DSW91] also speculates upon the connection to counterfactual logics, but does not develop the suggestion.

4 Reasoning

4.1 The problem

We suspect that the sound "logic" corresponding to any particular notion of preference is likely to be a weak one. For instance, the sequence of papers [DSW91, DW91, DW94] present various (related) definitions of preference, each of which is, in itself, far stronger than the technique of comparing of conditional expected utility. Yet the associated logics are quite limited. As Doyle, Shoham, and Wellman say in the conclusion of [DSW91]:

> "While the logic displays some intuitive properties ... some common and seemingly natural goal operations are not always valid.
>
> ...
>
> The numerous restrictions ... limit the applicability of the inference rules presented here."

Even among these inference rules, not all are (at least in our opinion) reasonable. For example, in the system of [DSW91], whenever φ logically entails ψ, then each of φ and ψ must be at least as desirable as the other.[9] In other words, their notion of relative desire cannot be used to distinguish between stronger or weaker assertions. This seems very unintuitive to us.

The truth seems to be that there are rather few "logical" laws governing preference which have strong and general intuitive support. The makes it difficult to develop a usefully rich logic for qualitative decision making. We are aware of two responses to this problem. The first approach is that taken by [Bou94, TP94a, TP94b]. These papers augment a rather weak underlying theory with some form of non-monotonic (and hence, unsound) reasoning. For example, [TP94a] are able to draw stronger conclusions by looking at what follows in preferred models that minimize the distinctions between the utilities of states. [Lou90] gives a general discussion and defense of the idea of non-monotonically reasoning about utilities.

Although the idea of using non-monotonic reasoning is surely a promising one, it seems too early to assess its success. One difficulty is that the choice of non-monotonic reasoning system used can appear rather arbitrary. For example, [TP94a] do not provide any extended justification for the definition they present, although there are clearly many alternatives that they could have used instead. Nor we aware of any specific proposal that has been applied to more than one or two examples.

The alternative to non-monotonic reasoning, that we are suggesting, is equally speculative. The idea is that instead of finding a logic for a single definition of preference or desirability, one should consider *all* the diverse sources of qualitative or semi-qualitative information one has—probabilistic independence, logics of likelihood, extreme probabilities, logics of preference and obligation, extreme utilities, independence assertions about utility and preference (the specific contribution of this paper), and more. Even quantitative information should be considered (so long as one is not asked for *all* of the numbers). Our conjecture is that together all these sources of information may enable quite sophisticated reasoning even though (in the absence of non-monotonic reasoning) this may not be the case for any one or two of them alone.

This paper is a step towards supporting this hypothesis. By considering in detail a formalism that allows one to state independencies of various types, we

[8] It might seem that we are exaggerating the connection to counterfactual logic, because semantics for counterfactual logics generally do not use probabilities. However, it is easy to show that standard counterfactual semantics are largely equivalent to certain well-known theories of qualitative probabilities (such as the κ-calculus [Pea93]).

[9] Note that it does not follow from this that their system collapses, because their notion of comparison is not necessarily transitive.

show that such information can support some useful inferences about preference. Nevertheless, in isolation such independence assertions are still not that powerful. Examining combinations of various pieces of information in the context of larger and more realistic problems remains important future work.

In the next section we present a small selection of sound reasoning patterns that take advantage of independence. These results are no more than suggestive of the usefulness of independence assertions in more realistic settings, but they do demonstrate that independence can be used to support some intuitive inferences that one might want to make about preferences. In fact, independence is often needed to ensure that these inferences are sound.

4.2 Some results

Suppose we prefer φ to $\neg\varphi$ and ψ to $\neg\psi$ (i.e., $\varphi \succ \neg\varphi$ and $\psi \succ \neg\psi$ using the semantics for $\succ$ given in Section 3.1). Would we prefer to have them both be true to having just one true, or to them both being false? At first glance one's response might be yes, this seems like a reasonable inference. Yet it is easy to construct counter-examples. Suppose, for instance, that Sue likes John and she also likes Fred. She might prefer to be married to John over not, and also prefer to be married to Fred over not. But at the same time might reasonably prefer to be married to neither over being married to both!

This leads to the obvious (and important) question of when it is in fact legitimate to assert that the combination of preferred goals is preferred. There are presumably many pieces of additional knowledge which could validate such reasoning. As our first result shows, utility independence can sometimes help.

Proposition 4.1 : *If $\varphi \succeq \neg\varphi$, $\psi \succeq \neg\psi$, and* either[10] *φ is utility independent of ψ or ψ is utility independent of φ, then $\varphi \wedge \psi \succeq \neg\varphi \wedge \neg\psi$*

Intuitively, if one believes that the direction of preference for (or against) φ, say, would not changed according to the value of ψ, then there is a limit to how undesirable their interaction can be. The result shows that getting two goods is preferable to getting neither *when independence holds*. In the example with Sue it is clear that utility independence does not hold.

"Monotonicity" of preferences is another very important pattern of reasoning. That is, when does $\varphi \wedge \pi \succeq \psi \wedge \pi$ follow from $\varphi \succeq \psi$, where π is another formula? This is actually quite a strong conclusion. One straightforward way of justifying it requires both utility and probabilistic independence.

[10] Utility and preferential independence are not symmetric.

Proposition 4.2: *If $\varphi \succeq \psi$, $\{\varphi, \psi\}$ is utility independent of π, and $\{\varphi, \psi\}$ is probabilistically independent of π,*[11] *then $\varphi \wedge \pi \succeq \psi \wedge \pi$.*

Both independencies are necessary, and utility independence cannot be replaced by preferential independence, if this result is to hold. On the other hand, there are different assumptions that lead to the same conclusion. For instance, suppose we assume that utilities are qualitative, in the sense of [TP94a]. (We omit a formal definition, but the basic idea is that utilities are ordinal ranks, in which maximization replaces addition.[12]) Then with qualitative utilities of this type the assumption of probabilistic independence in the previous result can be dropped.

Proposition 4.3 : *If $\varphi \succeq \psi$ and $\{\varphi, \psi\}$ is utility independent of π, then $\varphi \wedge \pi \succeq \psi \wedge \pi$ if utilities are qualitative.*

Additive independence can also be used in conjunction with other knowledge to obtain useful conclusions, as the next result illustrates. Suppose φ is preferred to ψ. Although it might seem intuitive at first that φ alone (i.e., $\varphi \wedge \neg\psi$) should be preferred to ψ alone (i.e., $\neg\varphi \wedge \psi$), a moment's reflection shows that this does not necessarily follow. For example, the news that one didn't win the state lottery (φ) is probably not as upsetting as learning that one's monthly paycheck has been canceled (ψ). But losing the lottery and receiving one's regular pay on time ($\varphi \wedge \neg\psi$) may well be inferior to winning the lottery but losing one's pay ($\neg\varphi \wedge \psi$). If we have additive independence (which may be plausible in this example) and that the second event is less likely than the first (not true in this case), such situations cannot occur:

Proposition 4.4 : *If $\varphi \succeq \psi$, $\{\varphi, \psi\}$ is additively independent, and ψ is less probable φ (i.e., $Pr(\psi) \leq Pr(\varphi)$), then $\varphi \wedge \neg\psi \succeq \neg\varphi \wedge \psi$.*

5 Actions and Decisions

In the presentation of this paper we have ignored any explicit discussion of *actions* and decision-making. In principle, what one really wants to do is to consider a family of probability distributions (parameterized by possible actions). Preferential comparisons are usually (but not invariably) of interest because they relate to two or more possible courses of action one might take.

A good response to this concern is given by Jeffrey. As he notes, it is usually possible to treat actions simply as new propositions. Thus, we might have propositional

[11] By which we mean that, for any atom A over $\{\varphi, \psi\}$, $Pr(A|\pi) = Pr(A|\neg\pi) = Pr(A)$.

[12] An alternative but equivalent semantics considers standard utilities of the form $\pm(1/\epsilon)^k$, then considers the limit as $\epsilon \to 0$.

symbols *do-A, do-B, ...* that are interpreted as true if and only if the corresponding actions *A, B, ...* are being performed. In this fashion, it can be argued, one avoids the need for any special treatment of actions. The decision between A and B reduces to deciding if *do-A* $\succ$ *do-B*, and if we have a detailed domain theory this may be resolvable within the current framework.

Furthermore, we are advocating that (where possible) knowledge be used that is in the form of qualitative assertions that constrain, but by itself does not fully determine, probability distributions. Such qualitative knowledge may be sufficiently robust that it applies to all possible actions being considered.

However, these responses are incomplete. We believe that the most important extension of the current paper is to investigate the idea of merging the work in this paper with a rich model of action. We believe that this would not require any changes to the basic semantics of preference and independence assertions that we are proposing here. Nevertheless, there remain many details that need to be investigated in order to make the formalism more useful. For example, an approach that might be integrated with the current work is the idea of distinguishing between "controllable" and "uncontrollable" propositions (for instance, as in [Bou94]).

6 Conclusions

In this paper, we have argued that *independence* concepts for utility and preference provides a category of *qualitative* information that can be useful for decision making. This is made more plausible by the analogy with probabilistic independence. By combining Jeffrey's notion of conditional expected utility with definitions from multi-attribute decision theory, we have given formal definitions that allow independence concepts to be used in a very general fashion. Our results show that these concepts can indeed be useful when reasoning about preferences.

Despite these results, one of the conclusions we reach is that the step from a numeric utility function to simple qualitative information about utilities is a large one. No single source or class of qualitative information seems to be that powerful in isolation. We suspect that a strong qualitative decision theory will need to take advantage of many diverse classes of information. Knowledge about independence is one such class, and should not be overlooked.

References

[BG95] F. Bacchus and A. Grove. Graphical models of preference and utility. In *Proceedings 11th Conference on Uncertainty in Artificial Intelligence (UAI 95)*, pages 3–19. Morgan Kaufmann, 1995.

[Bou94] C. Boutilier. Towards a logic of qualitative decision theory. In *Proc. Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR '94)*, pages 75–86, 1994.

[DSW91] J. Doyle, Y. Shoham, and M. P. Wellman. A logic of relative desire (preliminary report). In *Proc. 6th International Symposium on Methedologies for Intelligent Systems*, pages 16–31, 1991.

[DW91] J. Doyle and M. P. Wellman. Preferential semantics for goals. In *Proc. 9th National Conference on Artificial Intelligence (AAAI '91)*, pages 698–703, 1991.

[DW94] J. Doyle and M. P. Wellman. Representing preferences as *ceteris paribus* comparatives. In *AAAI Spring Symposium on decision-theoretic planning*, pages 69–75, 1994.

[Fis82] P. C. Fishburn. *The Foundations of Expected Utility*. Reidel, Dordrecht, 1982.

[Fre88] S. French. *Decision Theory*. Ellis Horwood, Chichester, West Sussex, England, 1988.

[Gin87] M. L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, San Francisco, CA, 1987.

[GS88] P. Gärdenfors and N. Sahlin, editors. *Decision, Probabilility, and Utility: Selected Readings*. Cambridge University Press, Cambridge, 1988.

[Jef65] R. C. Jeffrey. *The logic of decision*. University of Chicago Press, 1965.

[KLST71] D. H. Krantz, R. D. Luce, P. Suppes, and A. Tversky. *Foundations of Measurement*. Academic Press, New York, 1971.

[KR76] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Wiley and Sons, New York, 1976.

[Lew73] D. Lewis. *Counterfactuals*. Blackwell, 1973.

[Lou90] R. Loui. Defeasible reasoning about utilities and decision trees. In H. Kyburg, R. Loui, and G. Carlson, editors, *Knowledge Representation and Defeasible Reasoning*, pages 345–359. Kluwer, 1990.

[Pea88] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[Pea89] Judea Pearl. Probabilistic semantics for nonmonotonic reasoning: A survey. In *Proc. First International Conference on Principles of Knowledge Representation and Reasoning (KR '89)*, pages 505–516, 1989.

[Pea93] J. Pearl. From conditional oughts to qualitative decision theory. In *Proceedings 9th*

Conference on Uncertainty in Artificial Intelligence (UAI 93), pages 12–20. Morgan Kaufmann, 1993. A version of this paper appeared in the 1993 AAAI Spring Symposium Reasoning about Mental States, under the title "A Calculus of Pragmatic Obligation".

[Sav54] L. J. Savage. *The Foundations of Statistics.* Dover, New York, 1954.

[TH96] R. H. Thomason and J. F. Horty. Nondeterministic action and dominance: Foundations for planning and qualitative decision. In *Proceedings of the Sixth Conference on Theoretical Aspects of Reasoning about Knowledge (TARK-96)*, pages 229–250, 1996.

[TP94a] S. Tan and J. Pearl. Qualitative decision theory. In *Proc. 12th National Conference on Artificial Intelligence (AAAI '94)*, pages 928–932, 1994.

[TP94b] S. Tan and J. Pearl. Specification and evaluation of preferences under uncertainty. In *Proc. Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR '94)*, pages 530–539, 1994.

[von51] G. H. von Wright. Deontic logic. *Mind*, 60:1–15, 1951.

[von72] G. H. von Wright. The logic of preference reconsidered. *Theory and Decision*, 3:140–167, 1972.