Independence and Qualitative Decision Theory*

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

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

Abstract

Probabilistic independence has proved to be a fundamental tool that can dramatically simplify the task of eliciting, representing, and computing with probabilities. We advance the position that notions of utility independence can serve similar functions when reasoning about preferences and utilities during decision making.

In this paper we first summarize existing results and definitions concerning various independence concepts that can be applied to utility functions. We then review the results of our previous work in the area.

1 Introduction

Clearly, many subfields of A.I. are concerned with *decision making*. And yet until fairly recently the insights from decision theory—in particular the formal apparatus of probability, utility, and the expected-utility paradigm—have played a rather small role in such work.

Planning research provides a good example. In place of utility as a means of specifying desirable outcomes, traditional planning has concentrated on the more specialized notion of *goals*. Goals are typically "all or nothing" conditions. In the simplest cases, states where all of the goals are satisfied effectively have positive utility, while all other states, including states where only a subset of the goals are satisfied, have zero utility. Hence, the right decision is always to execute actions that achieve all of the goals specified, which is exactly the kind of plan that traditional planning algorithms search for.

We can view the traditional notion of goals and generating plans (i.e., making decisions about actions) as a very simple form of qualitative decision theory: it employs a qualitative abstraction of numeric utilities and avoids explicit utility maximization. With this abstraction we do not need to (and indeed, cannot) reason about tradeoffs between goals. Although limited in many respects, the notion of goals is a useful one and interesting applications can be effectively modeled this way.

Goals in planning are intended to provide an interface mechanism with some system capable of acting on our behalf. We model the capabilities of the system by specifying the actions it can perform, and then specify what we want the system to do by giving it goals. A planner, which automates this particular form of qualitative decision theory, serves to translate our goals into instructions that the system can understand, e.g., a specific sequence of actions. Thus, goals provide a simplified mechanism for controlling a possibly complex system.

We can view any decision theory, qualitative or quantitative, in this way. One purpose of such theories is to provide a mechanism for separating the specification of what we want—via goals, preferences, utility assignments, etc.—from the specification of what needs to be done to best achieve our desires. The aim, of course, is to generate a specification of what needs to be done automatically, based on the preferences and desires we specify.

From this point of view a useful decision theory must include a convenient language for specifying desires. Two particular areas of concern are the quantity and the "naturalness" of the information required. Furthermore, once one has specified utilities or preferences somehow, there is then the question of whether there is a feasible computational approach to solving the so expressed decision problem.

Goals, as utilized in planning research, are limited. In general, we may want to control a system by specifying our desires in more flexible ways. We may have graded desires, or desires that conflict or otherwise interact with each other. In such cases we need a decision theory that is capable of reasoning about tradeoffs between various specified goals. With respect to expressive power, traditional decision theory, which uses numeric utilities and the principle of maximum expected utility (MEU), provides the standard.

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

*This is a position paper, constructed largely around a survey of results from two earlier papers of ours [BG95,BG96]. The extensive excepts from these papers are used with permission.

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. These difficulties have been one of the motivations for interest in qualitative theories of probability and utility (i.e., preference), e.g., [von72, Bou94, Pea93, TP94a, TP94b, DW91, DW94, DSW91].

In the following sections we will discuss two recent papers of ours [BG95, BG96], which have the common theme of trying to apply various *independence* concepts (from the field of *multi-attribute utility theory*) to decision making. We have two motivations for pursuing this approach. Our primary motivation is the analogy with effective probability modeling. Probability modeling faces very similar concerns to those mentioned above, and the notion of probabilistic independence, e.g., as manifested in the theory of Bayesian networks [Pea88, SP90], has proved to be of great importance in addressing these concerns. Hence it seems reasonable that notions of utility independence may prove to be similarly useful. This motivation is not new, e.g., Doyle and Wellman [DW92] have previously put forward arguments for the relevance of utility independence in ensuring modular specifications, i.e., in reducing the quantity and increasing the naturalness of the information required. Our secondary motivation comes from fact that the field of multi-attribute utility theory contains many interesting results ([KR76] is an excellent reference). So far as we are aware, the relevance of these results for artificial intelligence is for the most part an unexplored topic (although there are certainly exceptions; in particular see [DW91, DSW91, DW94, DW92]).

It should be understood that this paper contains no new technical results; its sole purpose is to bring to the attention of the reader some of the key ideas of utility independence, and to summarize some of the earlier results we have been able to obtain [BG95, BG96]. In Section 2, we present some of the key notions of utility independence. This section is largely excerpted from [BG95], and is presented in fair detail as it serves as essential background for our own results, and hopefully new results by the inspired reader. In Section 3, we summarize some of the results we have obtained on graphical models for utility functions. These results originally appeared in [BG95]. Graphical models have proved to be very useful for reasoning about probability, and we hope that future work can make this true for utilities as well. In Section 4, we discuss our approach for moving utility theory beyond simple attributes so that preferences and independencies can be asserted of logical formulas. The ability to deal with logical formulas in a theory of preferences would allow for an easier integration of knowledge of preferences with other types of knowledge. The approach we discuss was originally presented in [BG96]. Finally, we present some conclusions and possible directions for future work in Section 5.

2 Independence

In many domains of interest we are concerned with a state space that can be represented as a product space over some collection of attributes. Such spaces fit well with the use of propositional logics: each primitive proposition can be viewed as a binary-valued attribute and the points in the product space become truth assignments.

The size of such a state space grows exponentially with the number of attributes. If each point in the state space is explicitly assigned an individual utility value, we would encounter significant difficulties in eliciting and manipulating all of these values. But one can hope that, in natural problems, the utility assignments will exhibit structure. The field of *multi-attribute utility theory* [KR76] has studied a number of independence concepts for preference and utility. Independence allows us to structure the utility function so as to reduce the number of independent parameters we must specify, and can often simplify the computation of expected utility (the essential step in solving the standard decision problem).

There are many distinct notions of independence in multi-attribute utility theory, several of which we summarize here. First we introduce some notation.

We assume that $V = \{v_1, \ldots, v_n\}$ is a fixed set of n *variables*. Each variable v has a domain d_v of two or more elements. We will generally use lower case letters to denote variables and upper case letters to denote sets of variables. Where necessary, Greek letters will denote values for particular variables.

The set of states S consists of the set of points in the product space $\prod_{i=1}^{n} d_{v_i}$. Each $s \in S$ is thus a vector of n values, one value for every variable. Clearly the size of S is exponential in the number of variables.

If $X \subseteq V$ then $f(X)$ stands for some real valued function all of whose arguments are in X, i.e.,

$$f(X) : \prod_{v \in X} d_v \longrightarrow I\!R.$$

The general form of a utility function is $u(V)$, which can thus require exponentially many independent utility assessments.

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

$$p_1 \succeq_u p_2 \quad \textit{iff} \quad \sum_{s \in S} p_1(s)\, u(s) \geq \sum_{s \in S} p_2(s) u(s),$$

where p_1 and p_2 are two distributions over S. That is, we prefer p_1 to p_2 if p_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. In the decision theory literature, probability distributions are often called lotteries, and we often say that u induces a preference ordering over lotteries on S

In the development of decision theory, it is natural to take the preference relation as primitive. Any relation satisfying fairly weak rationality conditions (which we don't repeat here, but see, e.g., [Sav54, Fis82, Fre88]) corresponds to some utility function exactly as above. (Furthermore, the utility function characterizing a preference relation is unique, up to affine transformations.) This exact correspondence between preference and utility is one of the fundamental theorems of decision theory. In the following, whenever we talk about a preference over V we mean a preference over lotteries over $S = \prod_{v \in V} d_v$ satisfying the standard rationality postulates.

The first definition of independence we consider is *utility independence*. Intuitively, a set of attributes X is utility independent of everything else, if when we hold everything else fixed (i.e., the values of attributes $V-X$), the induced preference structure over X does not depend on the particular values that $V-X$ are fixed to. Given utility independence we can assert preferences over (lotteries on) X that hold *ceteris paribus*—i.e., all else being equal.

Definition 2.1 : Consider preference $\succeq$ over V, $X \subset V$, $Y = V-X$. Let $\tilde{\gamma}$ be any particular element of $\prod_{v \in Y} d_v$. That is, $\tilde{\gamma}$ is a particular assignment of values to the variables in Y. Every probability distribution p over $\prod_{v \in X} d_v$ corresponds to a distribution p^* on $S = \prod_{v \in V} d_v$ such that p^*'s marginal on X is p and p^*'s marginal on Y gives probability 1 to $\tilde{\gamma}$. We define the *conditional preference over X given $\tilde{\gamma}$*, $\succeq_{\tilde{\gamma}}$, to be the preference ordering such that

$$p \succeq_{\tilde{\gamma}} q \quad \text{iff} \quad p^* \succeq q^*,$$

where p and q are any two distributions over $\prod_{v \in X} d_v$. ∎

Definition 2.2: The set of attributes X is *utility independent* of $V-X$ when conditional preferences for lotteries on X do not depend on the particular value given to $V-X$. That is,

$$\left(\forall \gamma, \gamma' \in \prod_{v \in V-X} d_v \right) p \succeq_{\gamma} q \text{ iff } p \succeq_{\gamma'} q,$$

where p and q are any two distributions over $\prod_{v \in X} d_v$. ∎

Utility independence fails, for instance, if one has a preference reversal between two mixtures of the attributes X, when some attribute in $V-X$ is changed. Judgments of utility independence would 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.

Example 2.3: Say that there are only two attributes *health*, with values H and $\overline{H}$ (healthy and not healthy), and *wealth* with values W and $\overline{W}$ (wealthy and not wealthy). If the agent's utility function u is defined as $u(HW) = 5$, $u(H\overline{W}) = 2$, $u(\overline{H}W) = 1$, and $u(\overline{H}\overline{W}) = 0$, then it can be seen that for the agent *health* is utility independent of *wealth* and *wealth* is utility independent of *health*. Intuitively, no matter what the agent's wealth is fixed to, it will always prefer gambles that yield H with higher probability. That is, the agent's preference for being healthy is the same no matter if the agent is wealthy or not. The same can be said about its attitude towards being wealthy. ∎

Utility independence is known to have several strong implications. We list a few, using [KR76] as our source. First, utility independence is equivalent to the existence of a utility function with a special functional form:

Proposition 2.4: *X is utility independent of its complement in a preference structure $\succeq$ if and only if $\succeq$ corresponds to some utility function of the form:*

$$u_{\succeq}(V) = f(V-X) + g(V-X)h(X)$$

where g is positive.[1]

Thus we must assess three functions, but each has fewer than $|V|$ arguments. This may mean that there are far fewer independent numbers to learn and to store. Most of the interest in utility independence in standard decision theory concerns the case of *mutual utility independence* where every subset of variables is independent of its complement:

Proposition 2.5 : *Every subset of variables is independent of its complement in $\succeq$ if and only if there exists n functions $f_i(v_i)$ (i.e., each f_i depends on a single variable), such that either*

$$u_{\succeq}(X) = \prod_{i=1}^{n} f_i(v_i) + c$$

for some constant c, or

$$u_{\succeq}(X) = \sum_{i=1}^{n} f_i(v_i).$$[2]

This is an extremely strong conclusion, allowing enormous simplification. The precondition of the theorem might seem to require $O(2^n)$ utility independence conditions, but since utility independence satisfies various closure properties we do not need this many. There are in fact several sets of n independencies that suffice; see [KR76]. However, the n assertions that each attribute individually is independent of the rest are not sufficient. In this case, the result is weaker:

Proposition 2.6 : *If every variable is utility independent of the rest there is a function $f_i(v_i)$ for each variable, such that $u_{\succeq}(V)$ is a multilinear combination of the f_i's.*

Thus we must assess n functions as well as (potentially exponentially many) constants to capture the interactions

[1] It is also clearly possible to arrange f and g so that $h(X) = u_{\succeq}(X, \tilde{\gamma})$ where $\tilde{\gamma}$ is an arbitrary fixed assignment to $V-X$. The function $u_{\succeq}(X, \tilde{\gamma})$ is sometimes called a *conditional utility function.*

[2] It is more usual to express the f_i in terms of conditional utility functions and multiplicative constants. This representation is easy to derive, or see [KR76].

among the f_i's. This may still represent a net gain. We conjecture that this case might be worth studying in the context of artificial intelligence applications, and in particular for giving a better decision-theoretic account of "goals". (This is because it seems generally reasonable to suppose that any single "goal" will be utility independent of everything else.)

A much stronger form of independence is *additive independence*. This can be defined in several ways, but the most useful for us is:

Definition 2.7: Let $Z_1, ..., Z_k$ be a partition of V. $Z_1, \ldots, Z_k$ is additively independent (for $\succeq$) if, for any probability distributions p_1 and p_2 that have the same marginals on Z_i for all i, p_1 and p_2 are indifferent under $\succeq$, i.e., $p_1 \succeq p_2$ and $p_2 \succeq p_1$. ∎

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.

Example 2.8: Consider the utility function given in Example 2.3 involving *health* and *wealth*. As the previous example pointed out, *health* was utility independent of *wealth*. However *health* is *not* additively independent of *wealth*. Consider the two probability functions p_1 and p_2, where $p_1(HW) = p_1(H\overline{W}) = p_1(\overline{H}W) = p_1(\overline{H}\overline{W}) = 1/4$, and $p_2(H\overline{W}) = p_2(\overline{H}W) = 0$, $p_2(HW) = p_2(\overline{H}\overline{W}) = 1/2$. We have $p_1(H) = p_2(H) = 1/2$ and $p_1(W) = p_2(W) = 1/2$. That is, p_1 and p_2 have identical marginals over *health* and *wealth*. Yet the expected utility under p_1 is 2, while the expected utility under p_2 is $5/2$. This shows that there exists two distributions with the same marginals that are not indifferent under the given utility function. That is, *health* and *wealth* are not additively independent.

Intuitively, the agent prefers being both healthy and wealthy more than might be suggested by considering the two attributes separately. It thus displays a preference for probability distributions in which health and wealth are positively correlated. ∎

Proposition 2.9 : *$Z_1, \ldots, Z_k$ are additively independent for $\succeq$ iff $u_\succeq$ can be written as*

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

for some functions f_i.

Naturally, the most interesting case is where all variables are additively independent separately, so that we only need to find one single-argument function for each variable. In the rest of the paper, we will be interested in additive independence for a partition of V into two parts, $V = X \cup Y$, unless we say otherwise. It would seem reasonable that these are easier to reason with than independence assertions about arbitrary partitions.

Conditional versions of both additive and utility independence can be defined. The definitions require that the specified independence hold whenever some subset of variables are held fixed. For instance,

Definition 2.10 : X and Y are *conditionally additively independent* (CA-independent) given Z (X, Y, Z disjoint, $X \cup Y \cup Z = V$) iff, for any fixed value $\tilde{\gamma}$ of Z, X and Y are additively independent in the conditional preference structure over $X \cup Y$ given $\tilde{\gamma}$.

In this case, we write *CAI* (X, Z, Y). ∎

Proposition 2.11 : *X and Y are additively independent given Z iff $u_\succeq$ can be written in the form $f(X, Z) + f(Z, Y)$.*

3 Graphical Models

All of the propositions presented in the previous section were to demonstrate that a utility function satisfying various independence criteria can be decomposed into simpler and more modular functional forms. Thus independence allows us to reduce the problem of specifying a utility function over a large collection of attributes into a collection of smaller problems (i.e., specifying functions that encode our preferences over smaller sets of attributes.)

Representing and reasoning about a utility function's independencies then becomes an important subtask. If we can more naturally represent independencies and compute new independencies from old, then perhaps we could provide useful new tools for understanding and utilizing the structural properties of our utilities and preferences.

It is important to note that although we have defined the concepts of utility independence using quantitative utility functions, they are not limited to the quantitative case. For example, in some previous works the approach taken has been to replace numeric utility and probability functions by 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]. The above notions of utility independence can be applied *mutatis mutandis* to qualitative utility functions so defined, and as in the purely numeric case they continue to provide a useful means of specifying additional structure.

For reasoning about probabilities, *graphical models* which capture probabilistic independencies have proved to be a very useful tool. These models provide useful aids to intuition when constructing probabilistic models and can be used to directly support efficient techniques for computing probabilities. Inspired by this we explored in [BG95] the possibility of constructing graphical models for utility independencies.

Our results are preliminary. In particular, we are not yet able to suggest computational mechanisms that can directly utilize the models we arrived at. Nevertheless, we were able to demonstrate that non-trivial graphical models for

certain notions of utility independence do exist. We hope that future work can uncover additional models that can be usefully applied to help us deal with preferences and utilities.

Perhaps the major result of [BG95] was:

Theorem 3.1: *The set of CA-dependencies generated by any utility function has a perfect map.*

A *prefect* map is a graph in which the vertices represent the variables over which the utility function is defined, and in which vertex separation corresponds exactly to CA-independence. That is, we can separate the set of vertices X from the set of vertices Y by removing from the graph the set of vertices Z if and only if we have $CAI(X, Z, Y)$ in the utility function.

From this result and Proposition 2.11 it can be shown that we can read off a functional form for the utility function directly from its perfect map. In particular, we have

Theorem 3.2: *$G = (V, E)$ is a CA-independence map for a utility function u (i.e., all independencies suggested by vertex separation in the graph hold of u) if and only if u has an additive decomposition over the set of maximal cliques of G.*

One of the reasons why this result is interesting is that the functional form generated by CA-independencies is precisely the form that has often been assumed to hold of a utility function in work on computing MEU, e.g., [JJD94, DDP88, ST90]. Nevertheless, we concede that it remains far from clear to us whether CA-independence is the best or most natural independence notion to use. Additional graphical models that can capture and reason with other forms of utility independencies are needed. Such work may eventually lead to models that could have a direct impact on the computational issues of decision making.

4 Preferences over Formulas

Section 2 defined all of the notions of independence in a rather simple context, involving product spaces of attributes. 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*.

Most existing research in qualitative decision theory has concerned itself 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 are propositional logic formulas. The related area of deontic logic also supposes that one should reason about preference and obligation in a logical setting.

There are a number of good reasons to want to deal with logical formulas when reasoning about preferences, especially when doing qualitative reasoning. Preferences can sometimes be more naturally expressed using logical formulas; the more fine-grained alternative (i.e., dealing just with the individual attributes) can become clumsy, and in a certain sense is not as expressive. But perhaps most important is that we want to integrate our knowledge of preferences with the rest of our knowledge, much of which will be in some logical form.

The key issue we face when dealing with preferences over formulas is assigning semantics to such assertions. When dealing with a propositional language the atomic semantic entities are individual truth assignments, and it is natural to want to provide semantics to preference assertions in terms of preferences over these atomic entities. However, a formula corresponds to a *set* of truth assignments (i.e., those truth assignments for which the formula is true). For instance, when we assert that φ is preferred to $\neg\varphi$, we must find a way of mapping this assertion about sets of truth assignments to some assertion(s) about individual truth assignments.

Consider the case of two propositions p_1 and p_2. The assertion that p_1 is preferred to $\neg p_1$ says that, in some sense, the set of states $\{p_1 \wedge p_2, p_1 \wedge \neg p_2\}$ is preferred to the set $\{\neg p_1 \wedge p_2, \neg p_1 \wedge \neg p_2\}$. What, if anything does it say about preferences between individual states like $p_1 \wedge p_2$ and $\neg p_1 \wedge p_2$? A number of approaches have been taken to translate preference assertions over formulas into preferences over states. Both Doyle, Shoham, and Wellman [DSW91], and Tan and Pearl [TP94b] treat such assertions as specifying an implicit *ceteris paribus* condition. Roughly speaking, when considering the assertion that φ is preferred to $\neg\varphi$, they first partition the state space into sets: Each set contains all states in which the propositional variables not appearing in φ take on some fixed truth values. (In the above example, where φ is p_1, the sets would be $\{p_1 \wedge p_2, \neg p_1 \wedge p_2\}$ and $\{p_1 \wedge \neg p_2, \neg p_1 \wedge \neg p_2\}$, because p_2 does not appear in φ.) Intuitively, *within* each such set, we may say that "all else is equal". They then interpret "φ is preferred to $\neg\varphi$" as asserting that, in each set, all those states where φ holds are preferred to any state where $\neg\varphi$ holds. (In the example, $p_1 \wedge p_2$ would be preferred to $\neg p_1 \wedge p_2$, and $p_1 \wedge \neg p_2$ preferred to $\neg p_1 \wedge \neg p_2$, but there is no preference induced between $p_1 \wedge p_2$ and $\neg p_1 \wedge \neg p_2$.) Thus, the assertion is restricted to preferences between collections of states, but among these states the preference is universal.

In a sense the reliance of these works on a "universal" interpretation of preference forces them to tie preference assertion over formulas to implicit *ceteris paribus* conditions. To interpret "φ is preferred to $\neg\varphi$" as meaning that *all* states satisfying φ are preferred to all states satisfying $\neg\varphi$ is impossibly strong. The implicit *ceteris paribus* condition has the advantage of tempering such assertions by restricting the sets of states over which the universal preference holds.

Nevertheless, even with such tempering this approach remains problematic. One of the problems is that it becomes

very difficult to override preferences given more specific information. One cannot easily 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. Thomason and Horty [TH96] provide some additional criticisms of these semantics for preference.

Our approach is different. It builds on Jeffrey's proposal in [Jef65], which we refer to as *conditional expected utility*, to define the semantics of preference assertions over formulas. Conditional expected utility 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. Using U we then say that φ is preferred to $\neg\varphi$ if the collection of states satisfying φ has greater conditional expected utility than the collection of states satisfying $\neg\varphi$.

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. 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.

This notion of preference is very natural, but by itself is rather weak (in a sense, it is a direct opposite to the idea of "universal" semantics for preference). But we can build on it, by providing a simply yet general mechanism where by a variety of utility independence assertions about formulas can be stated. These assertions can be (but need not be) stated completely independently of assertions about preference.

We will not present the details here, but the basic intuition is as follows. Given an independence assertion, we first imagine a *new* smaller space in which, intuitively, the truth or falsity of each formula mentioned in the assertion is a new attribute. For instance, to assert that φ_1 is utility independent of $\{\varphi_2, \varphi_3\}$, one of the points in the constructed space might be $(\varphi_1 : \textit{true}, \varphi_2 : \textit{false}, \varphi_3 : \textit{true})$. Obviously, such a point also corresponds to a certain *set* of states in the original space (in this case the set of all states where φ_1 and φ_3 are indeed true, and φ_2 false). So we can now use conditional expected utility in induce a utility function over the new space. The point of this is that, since the formulas we are interested are just attributes in the new space, we are able to apply the standard independence notions from Section 2 directly.

Of course, we are interested in what an assertion in the constructed space says about the underlying set of states, since it is these we really care about. As we discuss in [BG96], each such assertion has the effect of imposing a collection of algebraic constraints on the original utility and probability functions.

In this way, we have the freedom to make arbitrary independence statements about arbitrary logical formulas. This is in contrast to the earlier proposals we discussed, where the implicit use of *ceteris paribus* (a form of independence) is invoked in a fairly rigid fashion. In particular, these proposal depend on the syntax (in particular, the choice of primitive propositions) to determine how *ceteris paribus* is interpreted; [DW91] call this the problem of *framing*.

It may seem strange to bring probabilities into the interpretation of statements concerning utility independence, but in fact it makes sense philosophically. As we noted in Section 2, utility independence can be used to assert preferences over a subset of the primitive propositions that hold given that all of the other propositions remain fixed. But when dealing with formulas 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 formalization of utility independence over formulas 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.[3]

Standard independence definitions defined over the primitive propositions 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.

Our proposal allows one to state rich independence assertions in a uniform context with preference assertions about arbitrary logical formulas. Thus one can develop a theory of sound rules of inference. Here is a very simple example of a deduction enabled by independence considerations, but which is invalid in general (i.e., it is not always true if

[3] 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]).

we adopt the semantics of standard decision theory without making the independence assertion). The results gives one case in which we can conclude exclusive preferences (i.e., the desirability of having one goal or condition but *not* the other), from a specified non-exclusive preference.

Proposition 4.1 : *If $\varphi \succeq \psi$, $\{\varphi, \psi\}$ is additively independent, and ψ is less probable φ (i.e., $Pr(\psi) \leq Pr(\varphi)$), then $\varphi \wedge \neg\psi \succeq \neg\varphi \wedge \psi$.*

Other examples are provided in [BG96]. However, the resulting logic remains quite weak. The truth seems to be that there are rather few "logical" laws governing preference which have strong and general intuitive support. (But having said this, we note that the situation is far *worse* if one seeks logical laws that do not include independence assertions as premises). This makes it difficult to develop a usefully rich logic for qualitative decision making.

Nevertheless, we believe that this approach, of using various utility concepts to bolster somewhat the very weak basic "logic" of preferences, to be a valuable one. An important direction for future research along these lines is to develop mechanisms for increasing the inferential power of the theory.

5 Conclusions and Future Work

Our main conclusion is that there still much work to be done before a useful theory of preferences and utility can be developed. Representing and reasoning with utility independencies is a promising tool that can help us in this task. As our work has shown, utility independence notions have structure that can be naturally represented and reasoned with. Knowledge about independence can ease the problems of elicitation, and strengthen the inferences that can be made from a collection of preference assertions. It is also plausible that knowledge of independence can be utilized to speed up expected-utility computations (although demonstrating this is mostly future work).

For the future, we also feel that more work is particularly required in the following areas:

- Mechanisms, graphical or otherwise, for representing and reasoning with utility independence assertions. Such mechanisms eventually need to be tied into mechanisms for reasoning about utilities and preferences in general.
- Methods for expanding the range of useful inferences that can be generated from a collection of preference statements.

For the second point, we can offer two possible directions. First, rather than trying to find a logic concerned solely with assertions of preference or desirability, the best approach might be to consider a theory that can deal in an integrated fashion with *all* the diverse sources of qualitative or semi-qualitative information one might have—probabilistic independence, logics of likelihood, extreme probabilities, logics of preference and obligation, extreme utilities, independence assertions about utility and preference, 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 this appears not be the case if one considers any one or two of them alone.

Second, an orthogonal approach is to use non-monotonic reasoning. Non-monotonic reasoning has been suggested before, and has been utilized in [Bou94, TP94a, TP94b]. In these papers, a rather weak underlying theory is augmented with some form of non-monotonic 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. It is quite feasible that non-monotonicity can be combined with independence assertions to allow us to weaken the non-monotonic assumption used (and thus lessen the unintended consequences) without loosing some of the more useful inferences.

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.

[BG96] F. Bacchus and A. Grove. Utility independence in a qualitative decision theory. In *Proc. 5th International Conference on Principles of Knowledge Representation and Reasoning (KR '96)*. Morgan Kaufmann, 1996.

[Bou94] C. Boutilier. Towards a logic of qualitative decision theory. In *Proc. 4th International Conference on Principles of Knowledge Representation and Reasoning (KR '94)*, pages 75–86, 1994.

[DDP88] R. Dechter, A. Dechter, and J. Pearl. Optimization in constraint networks. In R. M. Oliver and J. Q. Smith, editors, *Influence Diagrams Belief Nets and Decision Analysis*, pages 411–425. Wiley, 1988.

[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.

[DW92] J. Doyle and M. P. Wellman. Modular utility representation for decision-theoretic planning. In *Proc. 1st International Conference on Artificial Intelligence Planning Systems (AIPS-92)*, pages 236–242, 1992.

[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.

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

[JJD94] F. Jensen, F. V. Jensen, and S. Dittmer. From influence diagrams to junction trees. In *Proc. 10th Annual Conference on Uncertainty Artificial Intelligence*, 1994.

[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.

[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.

[SP90] G. Shafer and J. Pearl, editors. *Readings in Uncertain Reasoning*. Morgan Kaufmann, San Mateo, CA, 1990.

[ST90] R. D. Shachter and J. A. Tatman. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2):365–379, 1990.

[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. 4th International Conference on Principles of Knowledge Representation and Reasoning (KR '94)*, pages 530–539, 1994.

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