Reprinted from *Working Notes of the AAAI Spring Symposium on Decision-Theoretic Planning*, March, 1994, Steve Hanks, Stuart Russell, and Michael Wellman, editors.

Representing Preferences as *Ceteris Paribus* Comparatives

Jon Doyle
Laboratory for Computer Science
Massachusetts Institute of Technology
545 Technology Square
Cambridge, MA 02139
doyle@lcs.mit.edu

Michael P. Wellman
Artificial Intelligence Laboratory
University of Michigan
1101 Beal Avenue
Ann Arbor, MI 48109-2110
wellman@engin.umich.edu

Abstract

Decision-theoretic preferences specify the relative desirability of all possible outcomes of alternative plans. In order to express general patterns of preference holding in a domain, we require a language that can refer directly to preferences over classes of outcomes as well as individuals. We present the basic concepts of a theory of meaning for such generic comparatives to facilitate their incremental capture and exploitation in automated reasoning systems. Our semantics lifts comparisons of individuals to comparisons of classes "other things being equal" by means of *contextual equivalences*, equivalence relations among individuals that vary with the context of application. We discuss implications of the theory for representing preference information.

1 Introduction

Decision-theoretic treatments of preferences represent the objectives of a decision maker by an ordering over the possible outcomes of available plans. In taking a decision-theoretic approach to planning, we view this ordering relation as an ideal, but cannot hope to completely and directly encode it in the planning system, as the domain of outcomes is combinatorially large or infinite, and the relevant preference criteria vary across problem instances. Therefore, in designing preference languages for decision-theoretic planning, we seek constructs for describing general *patterns* of preference that hold over *classes* of outcomes and situations. The result is a logic of preference affording flexible specification of objectives for planning, underpinned by a decision-theoretic semantics.

1.1 Preferences as comparatives

The theory presented here grows out of an effort to understand the relations between decision-theoretic preferences and problem-solving goals (Wellman and Doyle, 1991; Doyle *et al.*, 1991; Wellman and Doyle, 1992). In this work, we have found that the underlying notion of preference as a specification of relative desire supports the general statements about preferences we wish to express as well as our effort to relate the preference logic to decision theory. As our formal framework developed, however, we realized that the techniques useful for conveying preference information have no special ties to preference relations, so our treatment here presents the theory in somewhat greater generality. Nevertheless, specification of preferences and goals remain prime applications of our general approach to comparatives, and we use notation drawn from the domain of preferences to discuss abstract comparative relations among individuals.

1.2 Specifying comparatives and superlatives

Finding ways of formalizing knowledge that make specifications convenient for human informants and efficient for automated reasoning constitutes a central issue in knowledge representation. Human convenience usually means that the formalizations should stay close to common means of human expression. Formalizations that offer human conveniences sometimes also make for efficient reasoning, in that human communication places great value on compact specifications that directly entail the most important conclusions.

Comparatives and superlatives offer excellent examples of the tendency of humans to exploit succinct expressions of knowledge. Knowledge in many fields involves knowledge of comparative relationships, such as relative probability and desirability, relative height and weight, comparative attractiveness and dangerousness. Comparisons of individuals (Abby is taller than Bob, Carl is more handsome than Dan, Ella's graduating is likelier than Fred's, Guy's the best baritone) along these dimensions pose only routine problems; formalizers typically assume linear scales with which to measure degrees to which individuals exhibit these properties. These dimensional orderings induce preorderings (reflexive and transitive relations) on the appropriate sets of individuals, so that one assesses comparative statements applied to individuals by checking the agreement of the statements with the appropriate preorder, and assesses superlative statements applied to individuals by checking that the individuals hold maximal rank in the appropriate preorder.

But few people restrict their use of comparatives and superlatives to statements about individuals, as use of these constructs in reference to classes of individuals

(Danes are taller than Sicilians, wealth is better than poverty, California girls are beautiful, Texas is large) offers great efficiency in communication and informal reasoning. Unlike statements about individuals, however, these generic comparatives and superlatives pose many difficulties for formalization and have received a variety of treatments in artificial intelligence based on statistical theories (Loui, 1988; Pearl, 1988), nonmonotonic logics (McDermott and Doyle, 1980; McCarthy, 1980), and prototypical representatives of classes. While each of these approaches has its merits in different cases, we do not believe that they adequately capture the meaning of every common use of comparatives. To address this perceived inadequacy, we present below a theory of comparatives based on the notion of *ceteris paribus* lifting of the unproblematic preorderings of individuals to (relatively) unproblematic comparisons among classes of those individuals, in which one class is more whatever than another if individuals in the former class are more whatever than individuals in the latter, other things being equal. We formalize intuitive notions of "other things being equal" via context-dependent equivalence relations among individuals called *contextual equivalences*. One natural class of contextual equivalences follows the path of analytic geometry and multiattribute utility theory and factors the outcome space into the cartesian product of a number of smaller spaces. The factor spaces correspond to dimensions or "attributes," and "all else being equal" means varying one attribute while holding all others constant.

Space limitations prevent us from presenting more than the basic motivations and definitions of the theory, and also lead us to omit all proofs of theorems from this paper; see (Doyle and Wellman, forthcoming) for a full presentation, including all proofs. Section 2.1 summarizes the basic concepts of orderings of individuals. In Section 2.2, we extend individual orderings to propositions in the most natural way and show this simple "lifting" of comparisons to propositions inadequate, thus providing a formal motivation for considering *ceteris paribus* generic comparatives. Section 3.1 first introduces the basic notion of *contextual equivalence* that underlies the subsequent development, and Section 3.2 uses this to define the notion of *ceteris paribus* comparison. Some elementary results follow from these definitions alone, but the more useful results for reasoning about comparatives follow instead from special properties of contextual equivalences, which we sketch in Section 3.3 along with the relation between the theory presented here and the theory described in our prior papers (Wellman and Doyle, 1991; Doyle *et al.*, 1991). Section 4 discusses related work.

2 Formalization

2.1 Individual comparatives

We begin by considering a set Ω of distinct individuals or objects and a comparison relation $\succsim$ over Ω. While the mathematical theory of order identifies a variety of ordering relationships occurring in applications, we assume that the comparison relation $\succsim$ forms a total preorder, that is, a complete, reflexive, and transitive relation $\succsim$ over Ω. Total preorders partition the set of individuals into one or more subsets, with all individuals in a partition element equivalent with respect to $\succsim$, and all partition elements ordered into a strictly decreasing sequence by $\succsim$.

When $\omega \succsim \omega'$ we say that ω is *weakly greater than* ω', which means that the former outcome is ranked at least as highly as the latter. The strict order $\succ$ consists of the irreflexive part of $\succsim$, that is, $\omega \succ \omega'$ (ω is *strictly greater than* ω') if and only if (iff) $\omega \succsim \omega'$ but $\omega' \not\succsim \omega$. When both $\omega \succsim \omega'$ and $\omega' \succsim \omega$, we say the two outcomes are *equally great*, and write $\omega \sim \omega'$. When ω is weakly greater than any other ω' in Ω (or in some subset of Ω), we say that ω is *maximal* in Ω (or in the designated subset). If only ω is maximal, we may also say it is *greatest*.

Many comparative relations represent or underlie numerical representations of quantities or the degrees to which an individual possesses the quality in question, with these quantities or degrees expressed via *mensuration functions* $u : \Omega \rightarrow \mathbf{R}$, such that $\omega \succsim \omega'$ iff $u(\omega) \geq u(\omega')$. Since we assume nothing about $\succsim$ beyond ordinal comparisons, we may transform a given mensuration function u by any monotonically increasing function φ on the reals to obtain a new function $\varphi \circ u$ that also represents $\succsim$. We treat the qualitative order relation as the fundamental concept in our development, and consider numeric representations as derivative from qualitative orders. We thus ignore the special properties of some quantities (additivity in the case of mass, intensities in the case of preferences, etc.) that purely ordinal orderings omit.

2.2 Propositional comparatives

While comparisons of individuals serve many purposes, effective action in many domains also relies on comparisons of classes of individuals with regard to these same qualities. For example, a knitwear manufacturer selling clothes around the world may use the fact that Danes are taller than Sicilians to ship more large and tall size garments to Denmark than to Sicily, and an ambitious young corporate lawyer may use the knowledge that New York offers better opportunities than Denver in choosing where to live. These generic comparatives prove useful over and over again in spite of individual counterexamples.

As a first step toward understanding the meaning of such generic comparatives, we consider the idea of *lifting* comparisons of individuals to comparisons of sets of outcomes in the simplest possible manner.

By *proposition* we mean a set of individual objects, and we take the powerset $\mathcal{P}(\Omega)$ to be the set of all propositions. (For some purposes, such as modeling changes in background knowledge or probability theory, we restrict the set of propositions to a subset of $\mathcal{P}(\Omega)$, but we ignore such complications for the moment.) As usual, the set $\mathcal{P}(\Omega)$ forms a boolean lattice when we interpret the lattice operations meet, join, and complement as the set operations intersection, union, and complementation. We write p, q, r, etc., to indicate individual propositions, denote the complement $\Omega \setminus p$ as $\bar{p}$, and write pq to mean $p \cap q$. We say that an outcome ω *satisfies* a proposition p just in case $\omega \in p$. We sometimes treat propositions as their characteristic functions by defining

$$p(\omega) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \omega \in p \\ 0 & \text{otherwise.} \end{cases}$$

Thus ω satisfies p just in case $p(\omega) = 1$.

The simplest method of lifting comparisons of outcomes to comparisons of propositions says a proposition p is weakly greater than to a proposition q just in case every outcome in p is weakly greater than every outcome in q.

Definition 1 (Simple lifting) *For propositions $p, q \subseteq \Omega$, we define* lifted *weak, strict, and equal greatness comparisons by*

1. $p \succsim q$ *iff* $\omega \succsim \omega'$ *for each* $\omega \in p$ *and* $\omega' \in q$,
2. $p \succ q$ *iff* $p \succsim q$ *but* $q \not\succsim p$,
3. $p \sim q$ *iff* $p \succsim q$ *and* $q \succsim p$.

Clearly, if $p \sim q$, then $\omega \sim \omega'$ for all $\omega \in p$, $\omega' \in q$, and if $p \succ q$, then $\omega \succsim \omega'$ for each $\omega \in p$ and $\omega' \in q$ and $\omega \succ \omega'$ for some $\omega \in p$ and $\omega' \in q$.

We may observe the limitations of this method of lifting comparisons most easily by considering the resulting interpretation of superlatives. Under this interpretation, a proposition p is maximal just in case $p \succsim \bar{p}$, that is, if every individual satisfying (contained in) the proposition is weakly greater than every individual not satisfying it. However, this simple interpretation of superlatives breaks down when we consider multiple superlatives by preventing us from distinguishing all but the extreme cases. To see this, we first define the important concept of logical independence of propositions.

Definition 2 (Logical independence) *Two propositions are* (logically) independent *just in case they and their complements each contains individuals the other does not; formally, p and q are independent, written $p \perp q$, just in case none of pq, $p\bar{q}$, $q\bar{p}$, and $\bar{p}\bar{q}$ is empty. We say a set of propositions is* completely logically independent *if all basic boolean conjunctions of the propositions are nonempty (i.e., for $\{p, q, r\}$, we must have $pqr \neq \emptyset$, $pq\bar{r} \neq \emptyset$, etc.) We say p and q are* semi-independent *iff neither $p\bar{q}$ nor $\bar{p}q$ are empty.*

Note that $\emptyset$ and Ω are not independent or semi-independent of any proposition. With the notion of logical independence in hand, we see the inadequacy of the simplest interpretation of superlatives in the following result.

Theorem 1 (No comparatives) *Suppose $p_i \succsim \bar{p}_i$ for $0 \leq i \leq n$, with p_i and p_j logically semi-independent for $i \neq j$. Then*

1. *Any individual that satisfies every p_i is weakly greater than any individual that does not,*
2. *Any individual that satisfies at least one p_i is weakly greater than any individual that falsifies every p_i, and*
3. *All individuals that satisfy some p_i and falsify some p_j are equally great.*

This result means that the simplistic translation of individual comparatives to propositional comparatives distinguishes only three degrees of greatness: maximal, intermediate, and minimal. For example, if high paying jobs are better than low paying jobs and jobs with short hours are better than jobs with long hours, the interpretation implies that high-paying jobs with long hours are just as good as low-paying jobs with short hours. These implications thus preclude further specification of comparatives among the intermediate propositions, such as saying that one's personal preference is for high-paying jobs with long hours over low-paying jobs with short hours. Since one often wishes to stipulate such comparatives incrementally, this interpretation does not meet the demands of practical knowledge representation.

We conclude from Theorem 1 that to permit incremental specification of comparatives and superlatives, we must weaken the interpretation of propositional comparatives (or at least propositional superlatives) from comparing each individual satisfying the proposition with *every* individual not satisfying it. We instead compare individuals *only* when both are the same other things equal, and interpret superlatives to mean that that whenever two individuals are the same, other things equal, that one satisfying the superlative is preferred to the one not satisfying it.

3 Contextual comparisions

Formalizing the notion of comparing propositions *ceteris paribus* means formalizing when two individuals are the same, other things equal. We do this in two steps. First, we interpret "the same" by an equivalence relation $\equiv$ on Ω, so that $\omega \equiv \omega'$ means ω and ω' are the same, other things equal. Second, we interpret "other things equal" to mean that this equivalence relation may vary with the context under consideration. In the simplest case, this context involves only the propositions in question; but it might also involve other information relevant to the comparison.

3.1 Contextual equivalence

We formalize the contextual variation of equivalences with the notion of a *contextual equivalence*, an assignment of equivalence relations on individuals to each "context" described by propositions.

We write $\mathcal{E}(\Omega)$ to denote the set of all equivalence relations (reflexive, symmetric, and transitive relations) on Ω, and E, E', etc., for individual equivalence relations in $\mathcal{E}(\Omega)$. For every pair of equivalence relations E, E', the intersection $E \cap E'$ and the transitive closure of their union, which we write as $E \mathbin{\hat{\cup}} E'$, are also equivalence relations. Comparing equivalences with respect to set inclusion, we see that the set $\mathcal{E}(\Omega)$ contains both a least element $E_\perp = \mathrm{id}_\Omega$, that is, the identity relation on Ω, and a greatest element $E_\top = \Omega \times \Omega$, that is, the complete relation on Ω. In algebraic terms, $\mathcal{E}(\Omega)$ forms a distributive lattice under these operations, and a semilattice under each operation considered separately.

Definition 3 (Generated equivalences) *If $R \subseteq \Omega \times \Omega$ is a relation, we define R^*, the equivalence* generated *by R, to be the simultaneous reflexive, symmetric, and transitive closure of R.*

Thus R^* is the transitive closure of $E_\perp \cup R \cup R^{-1}$, and we may indicate equivalence relations by means of relations that mention only nonreflexive pairs and ignore symmetry.

In this paper we use sets of propositions to identify the contexts for propositional comparisons and define the notion of contextual equivalence as follows.

Definition 4 (Contextual equivalence) *A* contextual equivalence *on Ω is a function $\eta : \mathcal{P}(\mathcal{P}(\Omega)) \to \mathcal{E}(\Omega)$ assigning to each set of propositions $\{p, q, \ldots\}$ an equivalence relation $\eta(p, q, \ldots)$.*

If $\omega \; \eta(p, q, \ldots) \; \omega'$, we usually write $\omega \equiv \omega' \bmod_\eta p, q, \ldots$, and omit the mention of η when this does not cause confusion. Note that the equivalence assigned to a set of propositions does not depend on the order in which we might enumerate the elements of the set, or on repetitions in such an enumeration. We often also identify singleton contexts with their elements by calling $\eta(p)$ the equivalence assigned to p (when properly speaking it is the equivalence assigned to $\{p\}$).

3.2 Comparison *ceteris paribus*

We define propositional comparatives as weak greatness *ceteris paribus* with respect to a contextual equivalence as follows.

Definition 5 (Comparative greatness) *We say that p is* weakly greater than *q, written $p \trianglerighteq q$ and read briefly as "p over q," iff $\omega \succsim \omega'$ whenever*

1. *$\omega \in p\bar{q}$,*
2. *$\omega' \in \bar{p}q$, and*
3. *$\omega \equiv \omega' \bmod p\bar{q}, \bar{p}q$.*

We say that p is strictly greater than *q, written $p \triangleright q$, iff $p \trianglerighteq q$ but not $q \trianglerighteq p$. We say that p and q are* equally great*, written $p \bowtie q$, iff $p \trianglerighteq q \trianglerighteq p$.*

This definition compares individuals in $p\bar{q}$ and $\bar{p}q$ rather than individuals in p and q for the following reason. If both individuals under scrutiny are contained in p, then the only relevant basis for comparison among them is whether they satisfy q, which is not a comparison between p and q. Similarly, if both individuals are contained in q, the comparison concerns only p rather than both p and q. Taking away these cases leaves the definition stated above.

The *ceteris paribus* condition that individuals be compared with respect to otherwise equivalent properties serves two purposes. First, the reference to context allows us to avoid the unrealistic assertion that any individuals satisfying the propositions stand in the indicated relation of greatness. And second, by quantifying over these contexts, we are permitted to consider the comparisons in particular situations, where something is known about the individuals involved.

Though we will later consider contextual equivalences exhibiting interesting structure, we may first obtain some elementary results about comparative greatness strictly from the logical form of the definition, independent of any conditions on the contextual equivalence employed. The first two of these show that comparatives may be contraposed, and that comparatives are reflexive relationships in a very general sense.

Theorem 2 (Contraposition) *If $pr \trianglerighteq qr$, then $\bar{q}r \trianglerighteq \bar{p}r$.*

In particular, if $p \trianglerighteq q$, then $\bar{q} \trianglerighteq \bar{p}$.

Theorem 3 (Reflexivity) *If $p \subseteq p'$, then $p \bowtie p'$, $p \ntriangleright p'$, and $p' \ntriangleright p$.*

This means that comparative greatness does not distinguish propositions from stronger or weaker conditions (including Ω and $\emptyset$), so the interesting cases of comparatives all concern relations among logically semi-independent conditions. In particular, every proposition p and $\emptyset$ are equally great, as are p and Ω, that is, $\emptyset \bowtie p \bowtie \Omega$ for all $p \subseteq \Omega$.

Although we assume individual comparisons are transitive, comparative greatness need not be transitive in general, as seen in the following example.

Example 1 (Intransitivity) *Suppose $\Omega = \{\omega, \omega'\}$, $\omega \succ \omega'$, and $\omega \equiv \omega' \bmod \{\omega'\}$. Then $\{\omega'\} \ntrianglerighteq \{\omega\}$, even though (by Theorem 3) $\{\omega'\} \trianglerighteq \emptyset \trianglerighteq \{\omega\}$.*

Nevertheless, comparatives are transitive in a number of cases. The following theorem captures the trivial cases.

Theorem 4 (Trivial transitivity) *If $p \trianglerighteq q \trianglerighteq r$ and either q is p or r, or r is p, then $p \trianglerighteq r$.*

To observe a practical example of the difficulties involved in piecemeal specifications of preference information, consider the following.

Example 2 (Combining preferences) *Suppose that the global space of individuals has 5 elements, which we may think of as* `tall` *&* `thin`, `tall` *&* `trim`, `medium` *&* `thin`, `medium` *&* `trim`, *and* `short` *&* `fat`. *That is, we think of the individuals in terms of height and weight attributes, each of which has three values; the height may be* `tall`, `medium`, *or* `short`, *and the weight may be* `thin`, `trim`, *or* `fat`. *The space of individuals then reflects that individuals are* `short` *iff* `fat`. *Suppose further one specifies preferences over each of these attributes separately;* `tall` $\trianglerighteq$ `medium` $\trianglerighteq$ `short` *(taller is better), and* `thin` $\trianglerighteq$ `trim` $\trianglerighteq$ `fat` *(thinner is better). These two specifications entail the preferences* `tall` *&* `thin` $\succsim$ `tall` *&* `trim` *and* `medium` *&* `thin` $\succsim$ `medium` *&* `trim`, *but entail no preference relating* `short` *&* `fat` *to the other individuals.*

Thus if we want conjunctions of maximally great propositions (tallness and thinness) to also be maximally great, we must ensure that there are enough individuals (real or fictitious) to enable transitivity to work through individuals (in this case, transitivity would easily prefer `tall` & `trim` to `short` & `trim` to `short` & `fat`).

As a first application of the notion of comparative greatness, we improve the simple propositional lifting interpretation of propositional superlatives.

Definition 6 (Superlative greatness) *We say that p is* weakly maximal *(or* weakly maximally great*), and write $\trianglerighteq(p)$, just in case $p \trianglerighteq \bar{p}$. We say that p is* strictly maximal *(or* strictly maximally great*), and write* $\rhd(p)$, *just in case* $p \rhd \bar{p}$. *If* $p \bowtie \bar{p}$, *we call* p comparatively neutral *and write* $\bowtie(p)$.

Clearly, if $\unrhd(p)$, then not $\rhd(\bar{p})$, and the extremal propositions $\emptyset$ and Ω are weakly maximal but not strictly maximal.

One may use this notion of comparatives to investigate principles for reasoning with superlatives. For example, some forms of reasoning decompose propositional superlatives into sets of logically related propositions (for example, sets of propositions yielding the original one by conjunction or disjunction, as in subgoaling) and treat these new propositions as superlatives. We may ask whether such operations are sound with respect to our semantics. In fact, the semantics reveals that these operations are not always valid, as seen in the following example.

Example 3 (Nondecomposable superlatives) *Suppose* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $p = \{\omega_1, \omega_2\}$, $q = \{\omega_1, \omega_3\}$, $\eta(p) = \{(\omega_1\ \omega_3), (\omega_2\ \omega_4)\}^*$, $\eta(q) = \{(\omega_1\ \omega_2), (\omega_3\ \omega_4)\}^*$, *and* $\eta(p,q) = \eta(p) \cup \eta(q) = E_\top$. *If we also have* $\omega_1 \succ \omega_4 \succ \omega_2 \succ \omega_3$, *then* $\unrhd(pq)$ *and* $\unrhd(p \cup q)$ *but neither* $\unrhd(p)$ *nor* $\unrhd(q)$ *holds.*

More intuitively, one does not expect the tallest smartest Patagonians to also be the tallest Patagonians or the smartest Patagonians; indeed, the phrase "tallest smartest Patagonians" exhibits an essential ambiguity that alerts one to difficulties at the start. On the other hand, someone searching for the tallest smartest Patagonians could do worse than to begin by examining the tallest and the smartest Patagonians.

Intuitively, decomposing superlatives into conjunctions and disjunctions of putative superlatives need not always produce *bona fide* superlatives because the decomposition propositions may have undesirable properties ("side-effects") in addition to their relation to the compound superlative. In general, comparatives over complex propositions tell us little about comparatives over their constituent parts.

We do not mean to suggest that reasoners stop using useful manipulations like conjunctive and disjunctive decompositions of superlatives just because these operations can be unsound. We only mean to point out that if a reasoner uses unsound operations, then either it risks exhibiting judgements through its actions that conflict with its represented comparatives, or its reasoning introduces new assumptions that change the underlying comparison order. These possibilities deserve explicit recognition, and explicit treatment in some cases.

We have defined comparative greatness in terms of orderings over the set of all individuals, but in many cases the knowledge available to the agent or the agent's very constitution rules out the existence of some of these individuals. That is, some logically possible individuals may be epistemically or constitutionally impossible. In such cases, we should not demand that the agent express comparisons over the irrelevant logically possible individuals, but only require a comparison order over the epistemically or constitutionally possible individuals.

To make comparative greatness more practical in this way, we restrict the previous definitions to a set $r \subseteq \Omega$ representing the epistemically or constitutionally possible individuals, and define *restricted* comparative greatness as follows.

Definition 7 (Restricted comparative greatness) *We say that* p *is* weakly greater than q when restricted to r, *written* $p \unrhd^r q$ *and read briefly as "p over q in r," iff* $\omega \succsim \omega'$ *whenever*

1. $\omega \in p\bar{q}r$,
2. $\omega' \in \bar{p}qr$, *and*
3. $\omega \equiv \omega' \bmod p\bar{q}, \bar{p}q$.

We say that p *is* strictly greater than q when restricted to r, *written* $p \rhd^r q$, *iff* $p \unrhd^r q$ *but not* $q \unrhd^r p$. *We say that* p *and* q *are* equally great when restricted to r, *written* $p \bowtie^r q$, *when* $p \unrhd^r q$ *and* $q \unrhd^r p$.

This definition characterizes restricted comparative greatness in exactly the same way as comparative greatness, except that one considers only individuals in r rather than Ω. In other words, $p \unrhd q$ just means $p \unrhd^\Omega q$. Note that the scope restriction only affects what individuals are considered, but not the equivalence relation used to compare them.

As with plain comparative greatness, several simple results follow purely from the logical form of the definition and the logical relations among the propositions involved. The first of these is that stronger restrictions preserve relative greatness.

Theorem 5 (Strengthening) *If* $r' \subseteq r$, *then* $p \unrhd^r q$ *implies* $p \unrhd^{r'} q$.

In contrast, $p \rhd^r q$ does not imply $p \rhd^{r'} q$ because the stronger restriction r' may exclude all individuals which witness the strict greatness of p with respect to q.

An immediate consequence of this is that plain comparative greatness is the same as comparative greatness in every restriction.

Corollary 6 (Arbitrary restrictions) $p \unrhd q$ *iff* $p \unrhd^r q$ *for every* r.

Restrictions incompatible with the propositions being related lead to trivial restricted comparative greatness relationships.

Theorem 7 (Incompatible restrictions) *If* $p \subseteq \bar{r}$ *or* $q \subseteq \bar{r}$, *then* $p \unrhd^r q$.

In particular, $p \unrhd^\emptyset q$ holds for all p, q, as does $\unrhd^p(p)$ and $\unrhd^{\bar{p}}(p)$.

Finally, superlatives can vary quite easily with restrictions to different contexts.

Theorem 8 (Relativity of superlative greatness) *If* $p \unrhd q$ *does not hold, then* $q \rhd^r p$ *for some* r.

In particular, if p is not weakly maximal, then $\bar{p}$ is sometimes strictly maximal, and if neither p nor $\bar{p}$ are weakly maximal, then one can find different restrictions making each strictly maximal.

3.3 Sketch of the further theory

In the full presentation of the theory (Doyle and Wellman, forthcoming), we investigate some important special sorts of contextual equivalences.

We start by considering the very special contextual equivalences induced by familiar sorts of multiattribute representations of individuals. We consider the lattice structure of contextual equivalences to identify the *atomic* equivalences underlying or "supporting" a given equivalence, and derive multiattribute representations from arbitrary contextual equivalences by considering attributes constructed from the atomic equivalence classes. With this construction in hand, we develop conditions under which a contextual equivalence corresponds to a multiattribute representation in the sense that it is the same as the contextual equivalence induced by its derived multiattribute representation.

The remainder of the theory starts by identifying more general classes of contextual equivalences and derives results from their structure alone. The first property studied, *additivity*, makes the equivalence relation corresponding to the union of two sets of propositions to be the transitive closure of the union of the equivalence relations corresponding to the original sets of propositions. Members of the second class, *supported* contextual equivalences, in effect assign equivalences consisting of transitive unions of the atomic equivalences underlying them. The third class, that of *separated* contextual equivalences, is defined in terms of the notions of *orthogonal* equivalences and *relevance* of equivalences to propositions. We say two propositions are weakly orthogonal just in case their assigned equivalences intersect in only the identity; and a proposition is relevant to another just in case its assigned equivalence identifies some individual in the other proposition with some individual in the complement of the other proposition. We then define separated contextual equivalences as ones in which propositions are weakly orthogonal iff irrelevant. This notion captures the idea that irrelevant propositions involve completely different "dimensions" along which outcomes may vary without affecting each other. We put these notions all together to define *cartesian* contextual equivalences as additive, supported, and separated contextual equivalences.

We derive various important principles for reasoning about generic preferences from these notions. We show that cartesian comparative preference satisfies a dominance principle, in that if p and q are orthogonal to r, $pr \trianglerighteq qr$, and $p\bar{r} \trianglerighteq q\bar{r}$, then $p \trianglerighteq q$. We relativize the notion of logical independence to the notion of *contextual independence*, and show that for orthogonal and contextually independent propositions, conjunctions and disjunctions of goals (superlatives) are themselves goals. For such propositions, we also derive a variety of transitivity results; for example, propositions preferred to goals are also goals.

Finally, we relate the formalization provided here with that developed in our previous treatments (Wellman and Doyle, 1991; Doyle *et al.*, 1991). The first version of the theory, presented in (Wellman and Doyle, 1991) and based on pure attributive representations of outcomes, provided a definition of preferential superlatives (i.e., goals) but not of preferential comparatives. The second version, presented in (Doyle *et al.*, 1991), generalized the definition to comparatives, but phrased everything in terms of a logical (boolean attribute) representation. In particular, the definition of equivalence used was based on a syntactic notion of support rather than taking the notion of equivalence relation as primitive. Moreover, the present treatment compares outcomes using only one equivalence relation ($\equiv \bmod p\bar{q}, \bar{p}q$) rather than two ($\equiv \bmod p\bar{q}$ and $\equiv \bmod \bar{p}q$). If we ignore the syntactic definition of equivalence in the earlier theories and simply use the current notation anachronistically, we can rephrase the previous definition of preferential comparatives as $p \trianglerighteq q$ iff $\omega \succsim \omega'$ whenever $\omega \in p\bar{q}$, $\omega' \in \bar{p}q$, and there was some outcome ω'' such that $\omega \equiv \omega'' \bmod p\bar{q}$ and $\omega' \equiv \omega'' \bmod \bar{p}q$. We prove that the old and new definitions subsume each other in some cases, but not all, at least when one assumes this revisionist definition to be accurate.

4 Related work

The problem of representing preferences and goals for decision-theoretic planning has only recently drawn attention from AI researchers. Haddawy and Hanks have proposed specific techniques for representing goals as predicates with specified ranges of utility (Haddawy and Hanks, 1990) and for incorporating temporal factors with prototypical deadline models (Haddawy and Hanks, 1992). We come to this problem from a somewhat different (complementary) perspective, aiming to define logical constructs that avoid commitments to the precise form of utility, at the expense of perhaps providing weaker preference information.

Boutilier has recently proposed qualitative preference relations based on the concept of preference in the most likely or "normal" worlds (Boutilier, 1993). This preference, "all else being normal" contrasts directly with our preference, "all else being equal". Although we are still exploring the relation between these two approaches, it appears that the former more directly captures the default status of preferences, whereas the latter can express patterns of preference that hold in all (not just the most likely) contexts.

For other very recent AI research on representing preferences and utility, see the contributions by Koenig and Simmons, Linden, Mantha, and Tan in these *Working Notes*.

Thirty years ago, G. H. von Wright (von Wright, 1963; von Wright, 1972) proposed a "logic of preference" founded on the notion of preference *ceteris paribus*. The definition we employed in our "logic of relative desire" (Doyle *et al.*, 1991) was very similar to that adopted in the logic of preference, although von Wright did not uncover or address the difficulties inherent in syntactic treatment of the "all else" being held equal. Without a more powerful notion of contextual equivalence, the logic fails to support any but the simplest inferences (basically, just those exhibited by von Wright).

When we apply our theory of comparatives to the decision-theoretic notion of preference, the *ceteris* *paribus* condition of Definition 6 is a form of what multiattribute utility theory calls *preferential independence* (Gorman, 1968; Keeney and Raiffa, 1976), which requires that preference for each attribute of outcomes be independent of the other attributes. The usual definition of preferential independence, however, does not allow strictness to vary as in Definition 6. Moreover, preferential independence of a single attribute is, for two-valued attributes, identical to the generally stronger property of *utility independence*, which requires that the cardinal utility (preference over prospects) be invariant.

Finally, we note that the semantics developed here formalize the methods in our previous work on decision-theoretic planning, which defined preference for a proposition by specifying a *positive qualitative influence* on utility (Wellman, 1990). The use of qualitative influences in that work suggests how to extend our framework to account for preferences over ordinally scaled quantities in addition to propositions.

Acknowledgments

We thank Craig Boutilier, Annette Ellis, Peter Haddawy, John Horty, David McAllester, Joseph Schatz, Yoav Shoham, and Richmond Thomason for valuable discussions. Jon Doyle is supported in part by the USAF Rome Laboratory and DARPA under contract F30602-91-C-0018, and by National Institutes of Health Grant No. R01 LM04493 from the National Library of Medicine. Michael Wellman is supported in part by the Air Force Office of Scientific Research under grant No. F49620-44-1-0027.

References

Boutilier, Craig, 1993. A modal characterization of defeasible deontic conditionals and conditional goals. In *AAAI Spring Symposium on Reasoning about Mental States*. 30–39.

Doyle, Jon; Shoham, Yoav; and Wellman, Michael P., 1991. A logic of relative desire (preliminary report). In Ras, Z. W. and Zemankova, M., editors, 1991, *Methodologies for Intelligent Systems, 6*, volume 542 of *Lecture Notes in Artificial Intelligence*, Berlin. Springer-Verlag. 16–31.

Doyle, Jon and Wellman, Michael P., forthcoming. Contextual equivalence and *ceteris paribus* comparatives. In preparation.

Gorman, W. M., 1968. The structure of utility functions. *Review of Economic Studies* 35:367–390.

Haddawy, Peter and Hanks, Steve, 1990. Issues in decision-theoretic planning: Symbolic goals and numeric utilities. In *Proceedings of the DARPA Workshop on Innovative Approaches to Planning, Scheduling, and Control*. 48–58.

Haddawy, Peter and Hanks, Steve, 1992. Representations for decision-theoretic planning: Utility functions for deadline goals. In Nebel, Bernhard; Rich, Charles; and Swartout, William, editors, 1992, *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*. 71–82.

Keeney, Ralph L. and Raiffa, Howard, 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. John Wiley and Sons, New York.

Loui, Ronald P., 1988. Computing reference classes. In Lemmer, John F. and Kanal, Laveen N., editors, 1988, *Uncertainty in Artificial Intelligence 2*. North-Holland. 273–289.

McCarthy, John, 1980. Circumscription — a form of non-monotonic reasoning. *Artificial Intelligence* 13(1):27–38.

McDermott, Drew and Doyle, Jon, 1980. Non-monotonic logic—I. *Artificial Intelligence* 13:41–72.

Pearl, Judea, 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

von Wright, Georg Henrik, 1963. *The Logic of Preference: An Essay*. Edinburgh University Press, Edinburgh.

von Wright, Georg Henrik, 1972. The logic of preference reconsidered. *Theory and Decision* 3:140–167. Reprinted in (von Wright, 1984).

von Wright, Georg Henrik, 1984. *Philosophical Logic: Philosophical Papers, Volume II*. Cornell University Press, Ithaca, NY.

Wellman, Michael P., 1990. *Formulation of Tradeoffs in Planning Under Uncertainty*. Pitman and Morgan Kaufmann.

Wellman, Michael P. and Doyle, Jon, 1991. Preferential semantics for goals. In *Proceedings of the National Conference on Artificial Intelligence*. 698–703.

Wellman, Michael P. and Doyle, Jon, 1992. Modular utility representation for decision-theoretic planning. In *Proceedings of the First International Conference on AI Planning Systems*.