In *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*, Seattle, WA, July 31 - August 4, 1994, pp. 928 - 933.

Technical Report
R-214
April, 1994

# Qualitative Decision Theory*

**Sek-Wah Tan** and **Judea Pearl**
Cognitive Systems Lab, Computer Science Department
University of California, Los Angeles, CA 90024
< *tan@cs.ucla.edu* > < *judea@cs.ucla.edu* >

## Abstract

We describe a framework for specifying conditional desires "desire $\alpha$ by $\epsilon$ degrees if $\beta$" and evaluating preference queries "would you prefer $\sigma_1$ over $\sigma_2$ given $\phi$" under uncertainty. We refine the semantics presented in (Tan & Pearl 1994) to allow conditional desires to be overridden by more specific desires in the database. Within this framework, we also enable consideration of surprising worlds having extreme desirability values and the determination of degrees of preference.

## Introduction

This paper describes a framework for specifying conditional desires "desire $\alpha$ if $\beta$" and evaluating preference queries "would you prefer $\sigma_1$ over $\sigma_2$ given $\phi$" under uncertainty. Consider an agent deciding whether she should carry an umbrella, given that she sees that the sky is cloudy. Naturally, she will have to consider the prospect of getting wet $\neg d$ (not dry), the possibility of rain $r$, the cloudiness of the sky $c$, and so on. Some of the beliefs that influence her decision may be expressed in conditional sentences such as: "if I have the umbrella, then I will be dry", $u \rightarrow d$; "if it rains and I do not have the umbrella, then I will be wet", $r \wedge \neg u \rightarrow \neg d$; and "typically if it is cloudy, it will rain", $c \rightarrow r$. She may also have preferences such as "I prefer to be dry", $d \succ \neg d$; and "I prefer not to carry an umbrella", $\neg u \succ u$. From the beliefs and preferences above, we should be able to infer that the agent will prefer to carry an umbrella if she observes that the sky is cloudy, assuming that being dry is more important to her than not carrying an umbrella.

The research reported in this paper concerns such qualitative decision making process. Our aim is to eventually equip an intelligent autonomous artificial agent with decision making capabilities based on two types of inputs: beliefs and preferences. Beliefs, some of which may be defeasible, will be specified by normality defaults such as "if you run across the freeway, then you are likely to die", written $run \rightarrow \neg alive$. Preferences will be encoded in conditional sentences such as "if it is morning, then I prefer coffee to tea", written $coffee \succ tea \mid morning$. Figure 1 shows a schematic of the program. Each normality default $\varphi_i \xrightarrow{\delta_i} \psi_i$ and preference sentence $\alpha_i \succ_{\epsilon_i} \beta_i \mid \gamma_i$ will be quantified by an integer $\delta_i$ or $\epsilon_i$ that indicates the *degree* of the corresponding belief or preference. A larger degree implies a stronger belief or preference. The program will also accept queries in the form of $(\phi, \sigma_1 \succ \sigma_2)$, which stands for "would you prefer $\sigma_1$ over $\sigma_2$ given $\phi$?" The output of the program is the degree $\epsilon$ to which the preference $\sigma_1 \succ \sigma_2$ holds in the context $\phi$.

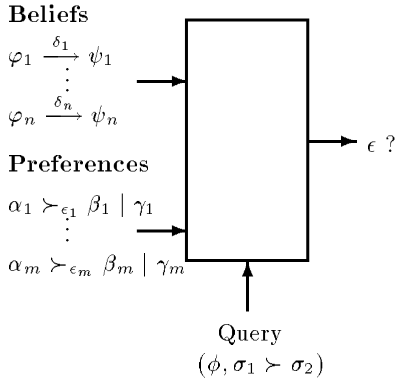


Figure 1: Schematic of the proposed system

The main obstacle in the way of constructing systems such as the one above is the unstructured nature of the input information. Decision theory, the traditional paradigm for rational decision making under uncertainty, requires a complete specification of a probability distribution and a utility function before reasoning can commence. Such complete specifications are impractical in complex tasks relying on commonsense knowledge, hence, one must find a way of transforming fragmented specification sentences, given in the form of normality and desirability expressions, into a coherent criterion for rational decision-making.

In previous work (Tan & Pearl 1994), preferences

*The research was partially supported by Air Force grant #AFOSR 90 0136, NSF grant #IRI-9200918, and Northrop-Rockwell Micro grant #93-124.

of the form $\alpha \succ \neg\alpha \mid \beta$ were given *ceteris paribum* (CP) semantics and interpreted as "$\alpha$ is preferred to $\neg\alpha$ other things being equal (ceteris paribum) in any $\beta$ world". Such conditional preferences are called *conditional desires* and written $D(\alpha|\beta)$. The problem with the CP semantics is that it does not handle *specificity* very well. In particular, if $D(\alpha)$ is a desire in the database, we cannot subsequently express a desire $D(\neg\alpha|\beta)$ for $\neg\alpha$ in a more specific situation $\beta$ (without modifying $D(\alpha)$ in the database as well). This is unsatisfactory as it is not uncommon for us to subscribe to some (default) set of desires (e.g., desire to be alive, to be rich, to be healthy) but subsequently qualify these desires (e.g., desire to die for some noble cause) for more specific situations. We would like to be able to handle specificity without having to examine or modify the desires that are already in the database.

In this paper we modify the CP semantics so that a conditional desire is allowed to override a less specific desire. We consider conditional desires of the form "if $\beta$ then $\alpha$ is desirable by degree $\epsilon$", written $D_\epsilon(\alpha|\beta)$, where $\alpha$ and $\beta$ are well-formed formulas (wffs) and $\epsilon$ is an integer. We will assume that the preferences of a reasoning agent may be represented by a *preference ranking* that is, by an integer-valued function on worlds which corresponds to an order-of-magnitude approximation of the agent's utility function. Conditional desires will be interpreted as constraints on *admissible* preference rankings. We will interpret a conditional desire $D(\alpha|\beta)$ as "$\alpha$ is preferred to $\neg\alpha$ ceteris paribum in any $\beta$ world if allowed by the other conditional desires in the preference database". While the CP semantics imposes *cp-constraints* between worlds that agree ceteris paribum, we have the additional requirement that a cp-constraint be not overridden by another cp-constraint that is due to a more specific conditional desire. A conditional desire is *more specific* than another if the former attempts to constrain a smaller set of worlds than the latter. To strengthen the system, we retain the principle of maximal indifference adopted in (Tan & Pearl 1994) and select from the set of admissible preference rankings the *most compact* rankings $\pi^+(\omega)$.

Another problem with the proposal in (Tan & Pearl 1994) is that preference queries are evaluated by comparing "believable" worlds. These are worlds that are ranked zero by the belief ranking, which is an integer-valued function that scores the "believability" of worlds. This excludes from consideration all worlds that are surprising (to any degree) even though some of the surprising worlds may have extreme positive or negative consequences. This is unsatisfactory as some extremely undesirable consequences (e.g., getting hit by a car), although unlikely, are not impossible, and some people would like to take such consequences into consideration. In this paper we weaken the notion of believability to include worlds that are ranked no more than some threshold $\delta$. We also extend the notion of preferential dominance and preferential entailment to allow for the strength of the preference to be determined and for the conclusions to be qualified by a degree of confidence.

In the next two sections we will describe the above extensions and improvements to the CP semantics. For the sake of brevity, we will not consider the semantics for normality defaults and will assume that we have a belief model that processes the input defaults and outputs a belief ranking. In the penultimate section, we compare related work and in the conclusion, we summarize the contributions of this paper.

# Conditional Desires

## Review and Notation

In this section we explain some notation and review some concepts that were introduced in (Tan & Pearl 1994). We consider conditional desires of the form $D_\epsilon(\alpha|\beta)$, where $\alpha$ and $\beta$ are wffs obtained from a finite set of atomic propositions $X = \{X_1, X_2, \ldots, X_n\}$ with the usual truth functionals $\wedge, \vee$, and $\neg$ and where $\epsilon$ is an integer. We will call $\alpha$ the desire, $\beta$ the condition, and $\epsilon$ the degree (or strength) of the conditional desire $D_\epsilon(\alpha|\beta)$. For simplicity we may write $D(\alpha|\beta)$ if the degree of the desire is not relevant to the discussion or if the degree is 1 (default value). When convenient we will also use the common form $\alpha \supset \beta$ instead of $\neg\alpha \vee \beta$.

A wff $\alpha$ *depends on* a proposition $X_i$ if all wffs that are logically equivalent to $\alpha$ contain the symbol $X_i$. The set of propositions that $\alpha$ depends on is represented by $S(\alpha)$. This set is referred to as the *support* of $\alpha$, written *support*($\alpha$), in (Doyle, Shoham, & Wellman 1991). The set of propositions that $\alpha$ does not depend on is represented by $\bar{S}(\alpha) = X \setminus S(\alpha)$. A world is simply a truth assignment on the set of atomic propositions. We will write $\omega = a\bar{b}$ to refer to the truth assignment that assigns *true* to $a$ and *false* to $b$. We say that two worlds *agree* on a proposition if they assign the same truth value to the proposition. Two worlds *agree* on a set of propositions if they agree on all the propositions in the set. We say that $\omega$ and $\nu$ are $S$-*equivalent*, written $\omega \sim_S \nu$, if $\omega$ and $\nu$ agree on the set $S \subseteq X$. We call $D(\alpha|\omega)$ a *specific* conditional desire if $\omega$ is a wff of the form $\bigwedge_1^n x_i$, where $x_i = X_i$ or $\neg X_i$. (As a convention we will use the same symbol $\omega$ to refer to the unique model of a wff $\omega$.) We assume that the preferences of the reasoning agent may be represented by a preference ranking $\pi$, which is an integer-valued function on the set of worlds $\Omega$. The preference rank of a world corresponds to an order-of-magnitude approximation of the utility associated with the world. The intended meaning of a ranking is that the world $\omega$ is no less preferred than the world $\nu$ if $\pi(\omega) \geq \pi(\nu)$. Given a non empty set of worlds $W$, we write $\pi_*(W)$ for $\min_{\omega \in W} \pi(\omega)$ and $\pi^*(W)$ for $\max_{\omega \in W} \pi(\omega)$. If $W$ is empty, we adopt the convention that $\pi_*(W) = \infty$ and $\pi^*(W) = -\infty$.

We associate with each conditional desire a set of worlds, called its *context*, which defines the worlds that the conditional desire constrains.

**Definition 1 (Context)** *Let $D(\alpha|\omega)$ be a specific conditional desire. The* **context** *of $D(\alpha|\omega)$, written $C(\alpha,\omega)$, is defined as*

$$C(\alpha,\omega) = \{\nu \mid \nu \sim_{\overline{S(\alpha)}} \omega\}. \tag{1}$$

*The context of a conditional desire $D(\alpha|\beta)$, written $C(\alpha,\beta)$, is defined to be $\cup_{\omega\models\beta} C(\alpha,\omega)$.*

Given a context $C$, we write $C_\gamma$ for $\{\nu \models \gamma \mid \nu \in C\}$ where $\gamma$ is a wff. We write $C(p)$ to represent the context of the conditional desire $p$.

## Specificity

In normal discourse, we have no difficulty accommodating general expressions of preferences that are subsequently qualified in more specific scenarios. For example, I desire to be alive $D(a)$, yet I am willing to die for some noble cause $D(\neg a|c)$. In the CP interpretation, this pair of desires would be inconsistent. In such a situation, we will usually allow $D(\neg a|c)$, which has a more specific condition, to override the unconditional desire $D(a)$. If a conditional desire attempts to constrain a subset of the worlds constrained by another desire, then we declare the former to be more specific than the latter. We define a conditional desire $p$ to be more specific than conditional desire $p'$ if the context of $p$ is a strict subset of the context of $p'$.

**Definition 2 (Specificity)** *Let $p$ and $p'$ be conditional desires. $p$ is* **more specific** *than $p'$, written $p \succ p'$, if $C(p) \subset C(p')$.*

Given this definition, one might wonder whether a simpler definition of specificity would suffice, one that considers only the conditions of the conditional desires. The simple proposal would be to declare a conditional desire $p = D(\alpha|\beta)$ to be more specific than another $p' = D(\alpha'|\beta')$ if the condition of $p$, $\beta$ implies the condition of $p'$, $\beta'$, that is, $\beta \supset \beta'$ and not vice versa. This simple proposal is not appropriate for our semantics because it declares the conditional desire $D(\alpha|\alpha)$ to be more specific than $D(\alpha)$ even though the two desires impose the same constraints. Although the simple definition is not suitable in general, it is sufficient when we restrict ourselves to a special type of conditional desires that we name simple.

**Definition 3 (Simple Desires)** *A conditional desire $D(\alpha|\beta)$ is* **simple** *if $\alpha$ and $\beta$ have disjoint support.*

**Theorem 1 (Simple Specificity)** *Let $p = D(\alpha|\beta)$ and $p' = D(\alpha'|\beta')$. If $p$ is simple and $\beta \supset \beta'$ but $\beta' \not\supset \beta$, then $p \succ p'$.*

Thus our definition of specificity coincides with the intuitive notion of when one conditional desire is more specific than another.

$$\begin{tikzcd}
ur \arrow[r] & \overline{u}r \\
u\overline{r} \arrow[u] & \overline{ur} \arrow[l] \arrow[u]
\end{tikzcd}$$

Figure 2: Applicable constraints in umbrella example.

## Admissible Rankings

In the CP semantics, *ceteris paribum* constraints are imposed independently by every conditional desire in the database.

**Definition 4 (CP-Constraints)** *$\nu \succ_\epsilon \nu'$ is a* **ceteris paribum constraint** *(cp-constraint) of a conditional desire $D_\epsilon(\alpha|\beta)$ if $\nu \in C_\alpha(\alpha,\omega)$ and $\nu' \in C_{\neg\alpha}(\alpha,\omega)$ for some $\omega \models \beta$. $c$ is a cp-constraint of a set $D$ if it is the cp-constraint for some $p \in D$.*

No consideration is given to possible conflicts among these constraints to accommodate specificity.

Let us consider how a cp-constraint may be "overridden". Consider a preference database consisting of two desires, the desire for good weather (no rain) $D(\neg r)$ and the desire to not carry the umbrella $D(\neg u)$. The cp-constraints of these desires are $\overline{ur} \succ u\overline{r} \succ ur$ and $\overline{ur} \succ \overline{u}r \succ ur$. Suppose that we would like to qualify our desire not to carry the umbrella by adding, to our database, the conditional desire to carry the umbrella if it is raining $D(u|r)$. The cp-constraint of $D(u|r)$ is $ur \succ \overline{u}r$, which is in direct conflict with the cp-constraint $\overline{u}r \succ ur$ of $D(\neg u)$. In this case we will like the cp-constraint $ur \succ \overline{u}r$ to override the cp-constraint $\overline{u}r \succ ur$ as $D(u|r)$ is more specific than $D(\neg u)$.

We say that two cp-constraints are in *competition* if they attempt to constrain the same worlds. For example the competing cp-constraints of $\omega \succ_\epsilon \nu$ are $\omega \succ_{\epsilon'} \nu$ and $\nu \succ_{\epsilon'} \omega$.

**Definition 5 (Admissible Rankings)** *A preference ranking $\pi$ is* **admissible** *with respect to a set of conditional desires $D$ if for all cp-constraints $\omega \succ_\epsilon \nu$ of $p \in D$ either*

$$\pi(\omega) \geq \pi(\nu) + \epsilon \tag{2}$$

*or there exists another sentence $p' \in D$ such that $p' \succ p$ and $p'$ induces a cp-constraint that competes with $\omega \succ_\epsilon \nu$.*

If there exists a ranking that is admissible with respect to a set of conditional desires $D$, then we say that $D$ is *consistent*. An example of an inconsistent set is $D = \{D(u), D(\neg u)\}$. Neither desire is more specific than the other and their cp-constraints are not overridden. Their cp-constraints are $u \succ \overline{u}$ and $\overline{u} \succ u$, respectively, and these imply $\pi(u) > \pi(\overline{u})$ and $\pi(\overline{u}) > \pi(u)$ since the default degree is 1. There is no ranking that can satisfy both inequalities simultaneously.

Table 1: Preference ranking in the umbrella example

| Worlds $\omega$ | Preference ranking $\pi^+(\omega)$ |
|---|---|
| $ur$ | $m+1$ |
| $\overline{u}r$ | $m$ |
| $u\overline{r}$ | $m+2$ |
| $\overline{u}\overline{r}$ | $m+3$ |

Table 2: Ranks in the fire example

| Worlds $\omega$ | Preferences $\pi(\omega)$ | Beliefs $\kappa(\omega)$ |
|---|---|---|
| $if$ | 1 | 1 |
| $\overline{i}f$ | $-1$ | 1 |
| $i\overline{f}$ | 0 | 0 |
| $\overline{i}\overline{f}$ | 1 | 0 |

In the umbrella example described above, we have the preference database $\{D(\neg r), D(\neg u), D(u|r)\}$. For a preference ranking to be admissible with respect to the database, it has to satisfy the cp-constraints that are not overridden. These cp-constraints are shown in figure 2 (an arrow $\omega \rightarrow \nu$ represents $\omega \succ \nu$, or $\pi(\omega) > \pi(\nu)$ since the default degree is 1). Here the cp-constraint of $D(\neg u)$, $\overline{u}r \succ ur$, is overridden by the cp-constraint $ur \succ \overline{u}r$ of $D(u|r)$ which is more specific.

In (Tan & Pearl 1994) we strengthened the semantics by adopting the principle of maximal indifference, which states that a reasoning agent is indifferent between two worlds unless a preference is explicitly communicated or can be inferred. We take the same approach here and select the most compact rankings from the set of admissible preference rankings. The reader is referred to (Tan & Pearl 1994) for a more complete discussion of the principle.

**Definition 6 (The $\pi^+$ Ranking)** *Let $D$ be a consistent set of conditional desires and let $\Pi$ be the set of rankings admissible with respect to $D$. A $\pi^+$ ranking is an admissible ranking that is* **most compact**, *that is*

$$\sum_{\omega,\nu \in \Omega} |\pi^+(\omega) - \pi^+(\nu)| \leq \sum_{\omega,\nu \in \Omega} |\pi(\omega) - \pi(\nu)| \qquad (3)$$

*for all $\pi \in \Pi$.*

Let us reconsider the umbrella example. The cp-constraints of the preference database (see figure 2) leads us to the most compact admissible preference ranking $\pi^+$. The ranks are shown in table 1 where $m$ is an integer.

## Preference Evaluation

### Normality Defaults

In evaluating preference queries, it is important that we be able to take into account the relative likelihoods of the worlds. The role of normality defaults in our proposal is to keep track of esoteric yet unlikely situations (just in case they become a reality) but not allow them to interfere with mundane decision making. For the sake of brevity, we will not describe in detail the treatment of normality defaults. We will instead assume that we have a belief model that accepts normality defaults representing qualitative expressions of beliefs and outputs a belief ranking $\kappa$, an integer-valued function on worlds which scores the "believability" of the worlds. An example of such a belief model can be found in (Goldszmidt 1992). Adopting Goldszmidt's convention, worlds with belief rank 0 are believable and an increasing rank indicates increasing surprise (or decreasing believability). We also assume that belief ranks are non negative. We will write $\kappa(\phi;\sigma_i)$ to represent the ranking that results after the execution of action $\sigma_i$ given context $\phi$. $\kappa^\delta(\phi;\sigma_i)$ will represent the set of $\delta$-believable worlds, namely, the set of worlds that have a $\kappa(\phi;\sigma_i)$ rank not greater than $\delta$.

In (Tan & Pearl 1994) consideration of the possible scenarios were restricted to the 0-believable worlds $\kappa^0$ while surprising ($\kappa > 0$) were completely ignored regardless of their preference ranks. This is unsatisfactory, as some of the surprising worlds may carry extreme positive or negative utilities. For example, consider fire insurance. Many people, despite believing fires ($f$) to be unlikely, still want to consider insuring their belongings ($i$). Table 2 shows a reasonable set of values for the beliefs and preferences associated with this example. If we were to concern ourselves only with the 0-believable worlds, then we would conclude that we prefer not to insure.

In this paper we extend preference consideration to $\delta$-believable worlds $\kappa^\delta$, where $\delta$ may be either a threshold specified by the user or an output value indicating a degree of confidence that qualifies the evaluation of the query. A preference query will be confirmed with confidence at least $\delta$ if it is confirmed with confidence $\delta - 1$ and also confirmed by considering $\delta$-believable worlds.

### Preferential Dominance

Preferential dominance (Tan & Pearl 1994) is a binary relation between sets of worlds which is derived from a preference ranking on worlds. Preferential dominance examines the three types of worlds that characterize the compared sets: the common possibilities, the additional possibilities, and the excluded possibilities. When considering whether we would prefer the set $W$ over the set $V$ (see figure 3), we imagine that the set $V$ represents the possibilities currently available to us and that the set $W$ represents the set of new possibilities. Let us consider the case when $W \subset V$. Since $W$ excludes some possibilities from $V$, we have to compare these excluded possibilities (in $V \setminus W$) with the new

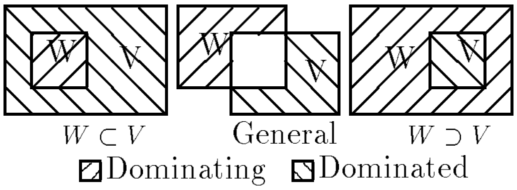


Figure 3: Interesting cases for $W \succ_\pi V$

Table 3: Ranks in the umbrella example

| Worlds $\omega$ | Preferences $\pi^+(\omega)$ | Beliefs $\kappa(\omega)$ |
|---|---|---|
| $ucr$ | $m+1$ | 0 |
| $\overline{u}cr$ | $m$ | 0 |
| $uc\overline{r}$ | $m+2$ | 1 |
| $\overline{u}c\overline{r}$ | $m+3$ | 1 |

possibilities offered by $W$. If the excluded possibilities are ranked lower than those that remain then $W$ protects us from those excluded possibilities and we should prefer $W$ to $V$. In the case when $V \subset W$, $W$ provides more possibilities. If these additional possibilities (in $W \setminus V$) are ranked higher than the current possibilities, $W$ provides an opportunity for improvement over the situation in $V$ and again we should prefer $W$ to $V$. In the general case, if $W$ and $V$ have some possibilities in common, then these common possibilities (in $W \cap V$) can be disregarded. If the additional possibilities (in $W \setminus V$) are ranked higher than the excluded possibilities (in $V \setminus W$), then we will prefer $W$ to $V$. In figure 3, $W$ $\pi$-dominates $V$, written $W \succ_\pi V$, if the worlds in the dominating set are preferred over the worlds in the dominated set. We generalize the notion of preferential dominance to allow a preference query to be confirmed with a degree indicating the strength of the confirmation.

**Definition 7 (Preferential Dominance)** *Let $\pi$ be a preference ranking and let $W$ and $V$ be two subsets of $\Omega$. We say that $W$* **$\pi$-dominates** *$V$ by $\epsilon$, written $W \succ_\pi^\epsilon V$, if and only if one of the following holds:*

1. *$\epsilon = 0$ when $W = V$,*
2. *$\pi_*(W) \geq \pi^*(V \setminus W) + \epsilon$ when $W \subset V$,*
3. *$\pi_*(W \setminus V) \geq \pi^*(V) + \epsilon$ when $W \supset V$, or*
4. *$\pi_*(W \setminus V) \geq \pi^*(V \setminus W) + \epsilon$ otherwise.*

*We write $W \succ_\pi V$ if $W \succ_\pi^\epsilon V$ for some $\epsilon \geq 0$ but $V \not\succ_\pi^0 W$.*

The definition of preferential dominance in (Tan & Pearl 1994) corresponds to $\succ_\pi^1$ and is therefore slightly stronger than $\succ_\pi$.

To evaluate the preference query $(\phi, \sigma_1 \succ \sigma_2)$ with degree $\epsilon$ and confidence $\delta$, we compare the set of $i$-believable worlds, $i = 0, \ldots, \delta$, resulting from executing $\sigma_1$ given $\phi$ to those resulting from executing $\sigma_2$ given $\phi$, and test if the former *preferentially dominates* the latter by $\epsilon$ in all the most compact preference rankings.

**Definition 8 (Preferential Entailment)** *Let $D$ be a set of conditional desires and $\kappa$ be some belief ranking on $\Omega$. $\phi$* **preferentially entails** *$\sigma_1 \succ \sigma_2$ with degree $\epsilon \geq 0$ and confidence $\delta$ given $\langle D, \kappa\rangle$, written $\phi \mathrel{|\!\!\sim}_\delta (\sigma_1 \succ_\epsilon \sigma_2)$, if and only if*

$$\kappa^i(\phi; \sigma_1) \succ_{\pi^+}^\epsilon \kappa^i(\phi; \sigma_2)$$

*for all $\pi^+$ rankings of $D$ and all $i = 0, \ldots, \delta$. We say that a preference query $(\phi, \sigma_1 \succ \sigma_2)$ is* **confirmed** *with degree $\epsilon$ and confidence $\delta$.*

We say that $\phi$ preferentially entails $\sigma_1 \succ_\epsilon \sigma_2$ with *absolute* confidence if $\phi \mathrel{|\!\!\sim}_\delta (\sigma_1 \succ_\epsilon \sigma_2)$ for all $\delta \geq 0$. We also write $\phi \mathrel{|\!\!\sim}_\delta (\sigma_1 \succ \sigma_2)$ if

$$\kappa^\delta(\phi; \sigma_1) \succ_{\pi^+} \kappa^\delta(\phi; \sigma_2)$$

for all $\pi^+$ rankings of $D$ and all $i = 0, \ldots, \delta$.

**Example**

Let us reconsider the umbrella story and the query "would you prefer to have the umbrella given that the sky is cloudy?", $(c; u \succ \neg u)$. We have the preference database $\{D(\neg r), D(\neg u), D(u|r)\}$. Let us assume that we have the defaults database $\{c \rightarrow r\}$. For this example we will adopt the belief model in (Goldszmidt & Pearl 1992; Pearl 1993). First we process the defaults database to get the resulting belief rankings $\kappa(\omega)$. Next, as in table 3, we list the possible worlds, given that the sky is cloudy, and obtain the belief ranking $\kappa(\omega)$ and the $\pi^+$ preference ranking (from table 1), where $m$ is some fixed integer. $\kappa^0(c; u) = \{ucr\}$ and has a minimum rank of $m+1$, while $\kappa^0(c; \neg u) = \{\overline{u}cr\}$ with a maximum rank of $m$. Therefore the preference query $(c; u \succ \neg u)$ is confirmed with degree 1 and confidence zero[1]. Unfortunately the preference query cannot be confirmed with absolute confidence.

## Comparison with Related Work

The assertability of conditional ought statements of the form "you ought to do $A$ if $C$" is considered in (Pearl 1993). The statement is interpreted as "if you observe, believe, or know $C$, then the expected utility resulting from doing $A$ is much higher than that resulting from not doing $A$". The treatment in (Pearl 1993) assumes, however, that a complete specification of a utility ranking on worlds is available and that the scale of the abstraction of preferences is commensurable with that of the abstraction of beliefs. Another problem is that the conclusions of the system are not invariant under a lateral shift of the utility ranking

[1]It is unfortunate that the phrase "confidence zero" conjures up the idea of a total lack of confidence which is definitely not the intended meaning of "being confirmed with confidence zero". The intended meaning is "considering only situations which are *serious* possibilities".

because the system endows worlds toward which the agent is indifferent with special status; for example, utility rankings $\pi_1$ and $\pi_2$, where $\pi_2(\omega) = \pi_1(\omega) + 1$, may admit different conclusions.

In (Boutilier 1994), expressions of conditional preferences of the form "$I(\alpha|\beta)$ - if $\beta$ then ideally $\alpha$" are given modal logic semantics in terms of a preference ordering on possible worlds. $I(\alpha|\beta)$ is interpreted as "in the most-preferred worlds where $\beta$ holds, $\alpha$ holds as well". This interpretation places constraints *only* on the most-preferred $\beta$-worlds, allowing only $\beta$-worlds that also satisfy $\alpha$ to have the same "rank". This contrasts with the CP semantics, which places constraints between pairs of worlds. In discussing the reasoning from preference expressions to actual preferences (which we here call preference queries), (Boutilier 1994) suggests that worlds could be assumed to be as preferred or as ideal as possible, parallelling the assumption made in computing the $\kappa^+$ belief ranking (Goldszmidt 1992) that worlds are as normal as possible. While it is intuitive to assume that worlds would gravitate towards normality because abnormality is a monopolar scale, it is not at all clear that worlds ought to be as preferred as possible since preference is a bipolar scale. This assumption of *maximal preference* can lead us to some very surprising conclusions though. Suppose that the only desire we have is to have bananas if we are alive $D(bananas \mid alive)$. The assumption will surprisingly deduce that our most desirable worlds include those where we are $\neg alive$. The $\pi^+$ rankings actually compacts the worlds away from the extremes thus minimizing unjustified preferences. It remains to be seen whether the $I$ operator corresponds closely with the common linguistic use of the word "ideally".

Ceteris paribum comparatives, relative desires and goal expressions have been considered in (Wellman & Doyle 1991; Doyle, Shoham, & Wellman 1991; Doyle & Wellman 1994). These accounts are similar to our semantics for unquantified unconditional desires. However, in their semantics (and also in Boutilier's system), preference constraints apply strictly to worlds satisfying the condition part of the preference statement. We believe that our interpretation captures a broader use of conditional desires in common discourse. For example, we often find an expression such as "if the light is off, we prefer to have the light on", $D(light \mid \neg light)$. This expression does not confine itself only to worlds satisfying $\neg light$. Instead, it actually compares worlds with *light* against those with $\neg light$. Although the desire *light* contradicts the condition, it is nevertheless an accepted way of specifying under what states of belief the preference would be invoked.

## Conclusion

This work refines the CP semantics (Tan & Pearl 1994) in three ways: it enables the handling of specificity of conditional desires, generalizes the notion of believability to allow consideration of surprising worlds, and extends preferential dominance so that the evaluation of preference queries may be qualified by an integer indicating the strength of the confirmation. The computational issues remain to be investigated, and further evaluation of the system needs to be done.

## Acknowledgments

We would like to thank the three anonymous reviewers for their constructive comments and suggestions. The first author is supported in part by a scholarship from the National Computer Board, Singapore.

## Proofs

**Lemma 1** *Let $p = D(\alpha|\beta)$ and $p' = D(\alpha'|\beta')$ be conditional desires. If $p$ is simple and $\beta \supset \beta'$, then $C(\alpha, \beta) \subseteq C(\alpha', \beta')$.*

**Proof:** Consider a world $\omega \in C(\alpha, \beta)$. Since $p$ is simple, we can find a $\nu \models \beta$ such that $\omega \in C(\alpha, \nu)$. This means that $\omega \sim_{\overline{S(\alpha)}} \nu$ (1). We construct the world $\nu'$ such that $\nu' \sim_{S(\alpha')} \nu$ (2) and $\nu' \sim_{\overline{S(\alpha')}} \omega$ (3). By definition, we have $\omega \in C(\alpha', \nu')$ (4).

Now we want to show that $\nu' \models \beta'$. We observe that $\nu' \sim_{\overline{S(\alpha)}} \nu$ by (1), (2) and (3). This implies that $\nu' \models \beta$ because $p$ is simple. $\nu' \models \beta'$ (5) follows from $\beta \supset \beta'$. Finally $\omega \in C(\alpha', \beta')$ by (4) and (5). ∎

**Lemma 2** *Let $p = D(\alpha|\beta)$ and $p' = D(\alpha'|\beta')$ be conditional desires. If $p$ is simple and $C(\alpha', \beta') \subseteq C(\alpha, \beta)$, then $\beta' \supset \beta$.*

**Proof:** Consider a world $\omega \models \beta'$. Clearly $\omega \in C(\alpha', \omega) \subseteq C(\alpha', \beta')$. By the premise of the lemma, $\omega \in C(\alpha, \nu)$ for some world $\nu \models \beta$. This implies that $\omega \sim_{\overline{S(\alpha)}} \nu$. Since $p$ is simple, $\omega \models \beta$ as well. ∎

**Theorem 1 (Simple Specificity)** Let $p = D(\alpha|\beta)$ and $p' = D(\alpha'|\beta')$. If $p$ is simple and $\beta \supset \beta'$ but $\beta' \not\supset \beta$, then $p \succ p'$.

**Proof:** The theorem follows from the previous two lemmas. ∎

## References

Boutilier, C. 1994. Toward a logic of qualitative decision theory. In Doyle, J.; Sandewall, E.; and Torasso, P., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94)*. Bonn, Germany: Morgan Kaufmann.

Doyle, J., and Wellman, M. P. 1994. Representing preferences as ceteris paribus comparatives. In Hanks, S.; Russell, S.; and Wellman, M., eds., *Working Notes of the AAAI Spring Symposium*, 69–75.

Doyle, J.; Shoham, Y.; and Wellman, M. P. 1991. The logic of relative desires. In *Sixth International Symposium on Methodologies for Intelligent Systems.*

Goldszmidt, M., and Pearl, J. 1992. Reasoning with qualitative probabilities can be tractable. In *Proceedings of the Eigth Conference on Uncertainty in Artificial Intelligence*, 112–120.

Goldszmidt, M. 1992. *Qualitative Probabilities: A Normative Framework for Commonsense Reasoning.* Ph.D. Dissertation, University of California Los Angeles, Cognitive Systems Lab., Los Angeles. Available as Technical Report (R-190).

Pearl, J. 1993. From conditional oughts to qualitative decision theory. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, 12–20.

Tan, S.-W., and Pearl, J. 1994. Specification and evaluation of preferences for planning under uncertainty. In Doyle, J.; Sandewall, E.; and Torasso, P., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference (KR94).* Bonn, Germany: Morgan Kaufmann.

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