# Risk Parameters for Utilitarian Desires (extended abstract)

**Leendert van der Torre**
Dept of AI, Vrije Universiteit
de Boelelaan 1081a, 1081 HV Amsterdam
The Netherlands
torre@cs.vu.nl

**Emil Weydert**
Max Planck Institute for Computer Science
Im Stadtwald, D-66123 Saarbrücken
Germany
weydert@mpi-sb.mpg.de

## Abstract

In qualitative decision-theoretic planning desires – qualitative abstractions of utility functions – are combined with defaults – qualitative abstractions of probability distributions – to calculate the expected utilities of actions. In this paper we consider Lang's framework of qualitative decision theory, in which utility functions are constructed from desires. Unfortunately there is no consensus about the desired logical properties of desires, in contrast to the case for defaults. To do justice to the wide variety of desires we define parameterized desires in an extension of Lang's framework. There are three parameters. The strength parameter encodes the importance of the desire, the lifting parameter encodes how to determine the utility of a set from the utilities of its elements, and the polarity parameter encodes the relation between gain of utility for rewards and loss of utility for violations. The parameters influence how desires interact, and they thus increase the control on the construction process of utility functions from desires.

## 1 Introduction

Classical decision theory [Luce and Raiffa, 1957; Jeffrey, 1983; Keeney and Raiffa, 1976] has been developed to describe and prescribe rational human decision making. However, due to so-called 'human irrationality', the description task is complicated so that its use may be restricted to decision making by artificial agents. For example, in decision-theoretic planning a robot receives our requirements or imperatives, tries to figure out the set of admissible utility functions and probability distributions, calculates the expected utilities and acts accordingly. However, a new problem arises for this application domain of decision theory. In planning it is assumed that we cannot completely impose our preferences and beliefs, because either we do not know them or it is computationally too expensive to elicitate and communicate them. These requirements are therefore as well *heuristic approximations* [Doyle and Wellman, 1991] as ways to *compactly* communicate our preferences and beliefs [Haddawy and Hanks, 1992] that only refer to *qualitative abstractions* of utility functions and probability distributions (the latter are sometimes called plausibilities). In qualitative decision theory these qualitative counterparts of preferences and beliefs are called desires and defaults. We summarize the terminology used in this paper in Table 1 below.

| utilities | | probabilities | |
|---|---|---|---|
| quantitative | qualitative | quantitative | qualitative |
| preference | desire | belief | default |

Table 1: Requirements in decision-theoretic planning

In this paper we propose a logic of utilitarian desires that builds on previous work of Boutilier [1994] and Lang [1996]. This logic is concerned with two problematic issues.

- First, as observed and discussed by Lang, the logic should not only characterize deductive relations between the desires – the logic of norms, imperatives and obligations called deontic logic for example also does so – but it should also determine the decision making process of the agent. As a consequence, Lang is more interested in the admissible utility functions than in the derivable desires. In other words, the semantics is more important than the syntactic or proof-theoretic counterpart.
- Secondly, not discussed or dealt with by either Boutilier or Lang, there are multiple intuitions about the logical properties of preferences and desires [Mullen, 1979; Pearl, 1993; Bacchus and Grove, 1996]. In other words, which desires can be derived intuitively sometimes depends on the meaning of the propositions. This multitude of intuitions hinders the effective use of desire specifications in a qualitative decision theory.

We give the robot's owner a tool to guide the robot's construction process of the intended utility functions by introducing several parameters.

**The strength parameter** encodes the importance of the desire,

- **The lifting parameter** determines how to construct the utility of a set from the utilities of its elements,
- **The polarity parameter** encodes the proportion between gain of utility for rewards and loss of utility for violation.

Decision theory explains the different intuitions about utilitarian desires and justifies our parameters. Rational agents base their decisions on the expected utility of their actions, i.e. they multiply the utility of the outcomes of possible actions by their probability and then choose the action that maximizes this expected utility. The intuitions differ due to the fact that utilities encode values as well as the agent's attitude towards risk, whereas probabilities only encode frequencies. They act *as if* they have an utility function, but they are not assumed to be aware of their compact values+risk representation. In classical decision theory, this unawareness is reflected by the contrived status of utility functions. To get some feeling for the different status of probabilities and utilities, consider the following two heuristics for requirements based on expected utilities. The first heuristic only considers the most likely states in the expected utility calculations, and the second heuristic only considers the most preferred states. The two heuristics are in an obvious way symmetric, but they have completely different consequences. The first heuristic cannot explain that people insure themselves for unlikely but grave events, see e.g. [Tan and Pearl, 1994a], and the second heuristic has the disadvantage that if the most preferred states are very unlikely, such as winning a lottery, then the requirement does not have an impact on the expected utilities and therefore not on the decisions.

With the parameters the risk component of each desire can be fit to the preference it encodes – we therefore call them risk parameters. The risk parameterization we propose for desires is not appropriate for defaults, though Boutilier's and Lang's logics are analogous to formalisms proposed for defaults, as we show in detail for Lang's framework and Weydert's framework for defaults. (They have as such been criticized by for example [Tan and Pearl, 1994b; Bacchus and Grove, 1996]). Our extension of the logic of utilitarian desires thus highlights a distinction between utilitarian desires and probabilistic defaults not found in the original proposals; we call it bipolarity.

## 2 The logic: explicit strengths

In this section we introduce the first parameter, that represents the strength $s$ of the desire. Weydert has introduced explicit strength parameters in $\models_{\geq 1}$ or in $\models_{>0}$ satisfaction, based on the following truth conditions for parameterized conditionals, where $u$ is a real-valued function on worlds.

$$u \models D_{\geq s}(a|b) \text{ if } \max_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w)$$
$$u \models D_{>s}(a|b) \text{ if } \max_{w \models a \wedge b} u(w) > s + \max_{w \models \neg a \wedge b} u(w)$$

There are no intuitive arguments supporting either one or the other because the two constraints are nearly identical. We have $u \models D_{>s}(a \mid b)$ if there is some small number $\epsilon$ such that $u \models D_{\geq s+\epsilon}(a \mid b)$. From the perspective of intuition, it is an arbitrary choice. However, there are technical distinctions. First, as we already remarked, $\models_{>0}$ determines a rational inference relation whereas $\models_{\geq 1}$ does not. Moreover, several constructions Weydert has investigated are easier defined in an extension of $\models_{\geq 1}$ satisfaction than in an extension of $\models_{>0}$ satisfaction. We therefore choose the former, abbreviating $D_{\geq s}(a|b)$ by $D_s(a|b)$. The results of this paper carry over to the other case.

### 2.1 The logic: the lifting problem

Consider the nonempty set of worlds that satisfy the proposition $p$ and an utility function $u$ that assigns utility to each of these worlds. What can we say about the utility of the set of worlds, i.e. the utility of $p$? This has been called the lifting problem (see e.g. [Thomason and Horty, 1996]), because the problem is how to lift a property of worlds to a property of sets of worlds.

Without knowing the probability of the individual worlds, the obvious choice is to consider the maximal or minimal utility of its elements. Let us call these operators $u_M(p)$ and $u_m(p)$, or $Mu(p)$ and $mu(p)$. $Mu(p)$ and $mu(p)$ are the poles of the set of utility values of the $p$ worlds, in the sense that for each world $w$ that satisfies $p$ we have that $Mu(p) \geq u(w) \geq mu(p)$. If we know that we are in a $p$ state, then assuming $Mu(p)$ is optimistic (the best case arises) and assuming $mu(p)$ is pessimistic (the worst case arises).

$$\begin{array}{lcl} Mu(p) & = & \max_{w \models p} u(w) \\ mu(p) & = & \min_{w \models p} u(w) \end{array}$$

$Mu(p)$ and $mu(p)$ can be used to define different types of constraints for desires (with strength $s$). The two poles can be compared in the following four ways, assuming there are $a_1$ and $a_2$ worlds.

$$\begin{array}{l}
u \models a_1 \succ_{mM:s} a_2 \\
\Leftrightarrow mu(a_1) \geq s + Mu(a_2) \\
\Leftrightarrow \min_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w) \\
u \models a_1 \succ_{MM:s} a_2 \\
\Leftrightarrow Mu(a_1) \geq s + Mu(a_2) \\
\Leftrightarrow \max_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w) \\
u \models a_1 \succ_{mm:s} a_2 \\
\Leftrightarrow mu(a_1) \geq s + mu(a_2) \\
\Leftrightarrow \min_{w \models a \wedge b} u(w) \geq s + \min_{w \models \neg a \wedge b} u(w) \\
u \models a_1 \succ_{Mm:s} a_2 \\
\Leftrightarrow Mu(a_1) \geq s + mu(a_2) \\
\Leftrightarrow \max_{w \models a \wedge b} u(w) \geq s + \min_{w \models \neg a \wedge b} u(w)
\end{array}$$

In Definition 1 below a desire $D(a|b)$ is defined as usual by $a \wedge b \succ \neg a \wedge b$. If either $a \wedge b$ or $\neg a \wedge b$ is inconsistent, i.e. if there are no worlds satisfying it, then we assume that the desire is vacuously true.

**Definition 1 (Logic of parametrized desires)** *A* (parametrized) desire *is defined by a pair of propositional formulas $a$ and $b$ together with a real $s > 0$ for strength and an index $l \in \{mM, MM, mm, Mm\}$ for lifting, and is denoted $D_{l:s}(a \mid b)$. A* (parameterized) desire specification $DS = \{D_{l_1:s_1}(a_1|b_1), \ldots, D_{l_n:s_n}(a_n|b_n)\}$ *is a finite set of parameterized desires. An* utility function *$u$, a map from $W$ to the reals $I\!R$, satisfies the desire $D_{l:s}(a|b)$, written as $u \models D_{l:s}(a|b)$, if and only if there are no $a \wedge b$ worlds, or there are no $\neg a \wedge b$ worlds, or according to the following truth conditions.*

$$\begin{array}{l}
u \models D_{mM:s}(a|b) \\
\Leftrightarrow mu(a \wedge b) \geq s + Mu(\neg a \wedge b) \\
\Leftrightarrow \min_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w) \\
u \models D_{MM:s}(a|b) \\
\Leftrightarrow Mu(a \wedge b) \geq s + Mu(\neg a \wedge b) \\
\Leftrightarrow \max_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w) \\
u \models D_{mm:s}(a|b) \\
\Leftrightarrow mu(a \wedge b) \geq s + mu(\neg a \wedge b) \\
\Leftrightarrow \min_{w \models a \wedge b} u(w) \geq s + \min_{w \models \neg a \wedge b} u(w) \\
u \models D_{Mm:s}(a|b) \\
\Leftrightarrow Mu(a \wedge b) \geq s + mu(\neg a \wedge b) \\
\Leftrightarrow \max_{w \models a \wedge b} u(w) \geq s + \min_{w \models \neg a \wedge b} u(w)
\end{array}$$

*An utility function $u$ satisfies the desire specification $DS$, written as $u \models DS$, if and only if it satisfies each desire in $DS$.*

The four types of desires directly imply the properties written below, in which we say that 'world $w_1$ is better than world $w_2$' if we have $u(w_1) > u(w_2)$.

| | |
|---|---|
| $u \models D_{mM:s}(a\|b)$ | each $a \wedge b$ world is better than all the $\neg a \wedge b$ worlds, |
| $u \models D_{MM:s}(a\|b)$ | the best $b$ worlds are $a$ worlds, or there are no $b$ worlds, |
| $u \models D_{mm:s}(a\|b)$ | the worst $b$ worlds are $\neg a$ worlds, or there are no $b$ worlds, |
| $u \models D_{Mm:s}(a\|b)$ | there is an $a \wedge b$ world that is better than a $\neg a \wedge b$ world, or there are no $b$ worlds. |

The following proposition shows the relations between the different types of desires.

**Proposition 2 (Relations between param. desires)** *We have the following relations between the parameterized desires based on the different values for the lifting parameter.*

- *if $u \models D_{mM:s}(a \mid b)$ then $u \models D_{MM:s}(a \mid b)$, $u \models D_{mm:s}(a|b)$ and $u \models D_{Mm:s}(a|b)$, and*
- *if $u \models D_{mM:s}(a \mid b)$, $u \models D_{MM:s}(a \mid b)$ or $u \models D_{mm:s}(a|b)$ then $u \models D_{Mm:s}(a|b)$,*
- *$u \models D_{MM:s}(a|b)$ does not imply $u \models D_{mm:s}(a|b)$ or vice versa.*

*These relations are represented in Figure 1 below.*

**Proof** *Follows directly from the fact that all truth conditions are universally quantified constraints on pairs of worlds, together with the fact that $Mu(a) \geq mu(a)$.*

$$\begin{tikzcd}
 & D_{MM:s}(a|b) \arrow[dr] & \\
D_{mM:s}(a|b) \arrow[ur] \arrow[dr] & & D_{Mm:s}(a|b) \\
 & D_{mm:s}(a|b) \arrow[ur] &
\end{tikzcd}$$

Figure 1: Relations between the four types of desires

Consider the additional assumption that the lifting parameters of all desires of the desire specification have the same value. In that case the lifting parameter is not a property of the individual desires but may be seen as a way we reason about desires. This is represented by indexing the satisfiability relation by the used lifting parameter, e.g. $\models_{mM}$, instead of the individual desires. In the following definition we say for the four lifting values $l$ that $DS$ is a $l$-conflict set if $DS$ is inconsistent with respect to $\models_l$.

**Definition 3 (Conflicts)** *A desire specification $DS$ (with only strength parameters) is an $l$-conflict set if there is no $u$ with $u \models D_{l:s}(a|b)$ for each $D_s(a|b) \in DS$. $DS$ is called conflict-free if it is not an $mM$-conflict set.*

We end this section with a brief discussion and illustration of the new types of desires. First, the desire $D_{mm:s}(a \,|\, b)$ is the dual of $D_{MM:s}(a \,|\, b)$ and has similar properties. As we already observed above, $D_{mm:s}(a \,|\, b)$ reflects a pessimistic view in the sense that it only considers the worst $b$ states, whereas $D_{MM:s}(a|b)$ only considers the best $b$ states.

Second, the desire $D_{mM:s}(a|b)$ induces a constraint on utility functions that is in the present setting too strong to be of much use, because it is rare that each $a \wedge b$ world is better than all $\neg a \wedge b$ worlds. For example, two desires 'to be healthy' $D_{s_1}(h \,|\, \top)$ and 'to be wealthy' $D_{s_2}(w \,|\, \top)$ are a $mM$-conflict set. Utility functions cannot satisfy the strong constraints if there are $w \wedge \neg h$ and $\neg w \wedge h$ worlds, because the first constraint prefers the first world to the second one and the second constraint vice versa. Moreover, a specificity set (there is a preference of no surgery over surgery, but this is inverse if surgery improves one's long term health [Bacchus and Grove, 1996]) is an $mM$-conflict set.

There is a set of examples for which the strong desires can be used, though. In other words, there are non-trivial conflict-free sets of desires. An example is the transitivity set discussed below, together with two other conflict free sets of desires. In this example we use the fact that the set of worlds can be restricted in the obvious way to all worlds which satisfy a set of formulas called the background knowledge – see [Lang, 1996] for details.

**Example 4 (Transitivity)** *Consider the following three desire specifications, together with the background knowledge $\neg(p \wedge c)$, $\neg(p \wedge h)$, $\neg(c \wedge h)$ and $\top \leftrightarrow (p \vee c \vee h)$. This background knowledge encodes that the three variables $p$, $c$ and $h$ are mutually exclusive and exhaustive. Hence, there are only $p \wedge \neg c \wedge \neg h$, $\neg p \wedge c \wedge \neg h$ and $\neg p \wedge \neg c \wedge h$ worlds in $W$. We also give the representation based on $\succ$ operators, because they are the most readible. CTD and ATD represent contrary-to-duty and according-to-duty examples extensively discussed in the logic of obligations, see e.g. [van der Torre and Tan, 1999].*

$$\begin{array}{lll}
\textit{TRANS} & D_{mM:1}(p \mid p \vee c) & p \succ_{mM:1} c \\
 & D_{mM:1}(c \mid c \vee h) & c \succ_{mM:1} h \\
\textit{CTD} & D_{mM:1}(p \mid p \vee c \vee h) & p \succ_{mM:1} c \vee h \\
 & D_{mM:1}(c \mid c \vee h) & c \succ_{mM:1} h \\
\textit{ATD} & D_{mM:1}(p \mid p \vee c) & p \succ_{mM:1} c \\
 & D_{mM:1}(p \vee c \mid p \vee c \vee h) & p \vee c \succ_{mM:1} h
\end{array}$$

*The three sets of constraints are equivalent. For all worlds $w_1, w_2, w_3$ such that $w_1 \models p$, $w_2 \models c$ and $w_3 \models h$ we have that $u(w_1) \geq 1 + u(w_2) \geq 2 + u(w_3)$.*

Finally, we consider the weakest desire $D_{Mm:s}(a|b)$. It seems to be too weak to be of any use, because there is nearly always an $a \wedge b$ world that is better than some $\neg a \wedge b$ world. However, some examples suggest that the three other constraints are too strong. One example is the marriage of Sue example of Bacchus and Grove [Bacchus and Grove, 1996].

**Example 5 (Marriage)** *Consider the desire specification DS that consists of the following three desires.*

| | |
|---|---|
| $D_1(j\|\top)$ | *Sue prefers to be married to John* |
| $D_1(f\|\top)$ | *Sue prefers to be married to Fred* |
| $D_1(\neg(j \wedge f)\|\top)$ | *Sue prefers to be married to neither* |

*DS is an mM-, MM- and mm-conflict set. For example, the desire specification*

$$\{D_{MM:1}(j|\top), D_{MM:1}(f|\top), D_{MM:1}(\neg(j \wedge f)|\top)\}$$

*is inconsistent, because there is not a single world that satisfies the materializations of all three desires ($j$, $f$ and $\neg(j \wedge f)$). In other words, each world violates at least one desire ($\neg j$, $\neg f$ or $j \wedge f$). However, DS is not an Mm-conflict set. An example of an utility function that satisfies the three desires $D_{Mm:1}(j|\top)$, $D_{Mm:1}(f|\top)$ and $D_{Mm:1}(\neg(j \wedge f)|\top)$ is*

$$u(w) = \begin{array}{rl} 0 & \text{if } w \models j \leftrightarrow \neg f \\ -1 & \text{if } w \models j \leftrightarrow f \end{array}$$

*We have $u \models D_{Mm:1}(j \mid \top)$ because $j \wedge \neg f$ worlds are better than $\neg j \wedge \neg f$ worlds, we have $u \models D_{Mm:1}(f \mid \top)$ because $\neg j \wedge f$ worlds are better than $\neg j \wedge \neg f$ worlds, and we have $u \models D_{Mm:1}(\neg(j \wedge f) \mid \top)$ because $j \leftrightarrow \neg f$ worlds are better than $j \wedge f$ worlds.*

A second example that is only consistent with the weakest constraint is the following desire specification.

**Example 6 (Fence and dog)** *Consider the desire specification DS that consists of the following three desires.*

| | |
|---|---|
| $D_1(\neg f\|\top)$ | *preference for no fence* |
| $D_1(f \mid d)$ | *preference for fence if there is a dog* |
| $D_1(d\|\top)$ | *preference for a dog* |

*DS is an mM-, MM- and mm-conflict set, but it is not an Mm-conflict set. An example of an utility function that satisfies the three desires $D_{Mm:1}(\neg f \mid \top)$, $D_{Mm:1}(f|d)$ and $D_{Mm:1}(d|\top)$ is*

$$u(w) = \begin{array}{rl} 0 & \text{if } w \models f \leftrightarrow d \\ -1 & \text{if } w \models f \leftrightarrow \neg d \end{array}$$

*We have $u \models D_{Mm:1}(\neg f \mid \top)$ because $\neg f \wedge \neg d$ worlds are better than $f \wedge \neg d$ worlds, we have $u \models D_{Mm:1}(f|d)$ because $f \wedge d$ worlds are better than $\neg f \wedge d$ worlds, and we have $u \models D_{Mm:1}(d|\top)$ because $f \wedge d$ worlds are better than $f \wedge \neg d$ worlds.*

Summarizing, there are desire specifications which can be analyzed with the strongest desires, and there are desire specifications which can only be analyzed with the weakest desires. However, most examples can more naturally be formalized with $D_{MM}$, i.e. with the semantics used in Boutilier's, Lang's and Weydert's frameworks. This will therefore be our standard representation.

## 2.2 The nonmonotonic construction

In this section we introduce our third parameter. We call it the polarity parameter $p$ and we express desires with polarity by $D^p_{l:s}(a|b)$. It is used in the local utility functions, i.e. in the construction of the distinguished utility functions. Consider a local utility function that not only considers loss of utility for violations, as in Lang's construction, but also gain of utility for rewards. That is, the real valued function $u$ is a local utility function of $D_{l:s}(a \mid b)$ - $u_{a|b}$ in Lang's notation - if there exists an $\alpha \geq 0$ (its utility loss) and a $\beta \geq 0$ (its utility gain) with $\alpha + \beta \geq s$ such that

$$u(w) = \begin{array}{rl} \beta & \text{if } w \models a \wedge b \\ 0 & \text{if } w \models \neg b \\ -\alpha & \text{if } w \models \neg a \wedge b \end{array}$$

For representational convenience we represent this utility function below by $u = u^{\beta}_{a \wedge b} + u^{-\alpha}_{\neg a \wedge b}$. The two reals $\beta$ and $-\alpha$ are the two poles of this local utility function, in the sense that for all worlds $w$ we have that $\beta \geq u(w) \geq -\alpha$. The polarity parameter is defined by $p = \frac{\alpha}{\alpha+\beta}$, and thus restricts the relative values of $\alpha$ and $\beta$. Obviously we have $0 \leq p \leq 1$. For example, mixed gain-loss desires with polarity 0.5 have their set of local utility functions $u$ defined for $\alpha \geq 0.5 \times s$ with $u = u^{\alpha}_{a \wedge b} + u^{-\alpha}_{\neg a \wedge b}$, i.e.

$$u(w) = \begin{array}{ll} \alpha & \text{if } w \models a \wedge b \\ 0 & \text{if } w \models \neg b \\ -\alpha & \text{if } w \models \neg a \wedge b \end{array}$$

If the polarity of a desire is 0 then we call the desire a gain desire, because its utility loss $\alpha$ is zero. Likewise, if its polarity is 1 then we call it a loss desire, because its utility gain $\beta$ is zero.

The philosophy of Lang's framework is to define the utility functions of a set of desires as a function of the utility functions of elements of this set; the latter are called their local utility functions. The same philosophy underlies multi-attribute utility theory with the use of additive independence [Keeney and Raiffa, 1976; Wellman and Doyle, 1992; Bacchus and Grove, 1996]. There are several different ways to represent this idea of defining the utility functions of a set of desires as a function from the utility functions of its elements. In this paper we follow a standard model preference semantics, similar to the one adopted by Weydert. Our reformulation of Lang's framework in standard model preference semantics has some advantages. Most importantly, in his definition it is unclear that there is a *set* of local utility functions associated with each desire, and that for the construction of the global utility function we have to choose elements from these sets. The representation in Definition 7 below also facilitates Proposition 8 afterwards. A second minor advantage is that logical notions such as inference relations are defined in the standard way.

Local and distinguished utility functions are defined in two steps. First the set of constructible utility functions is defined, represented by $CONS(DS)$, and thereafter the distinguished utility functions, represented by $U_J$ to refer to Jeffrey conditionalization. Due to this two step definition the distinguished utility functions are *not* simple additions of local utility functions. Instead, in Proposition 8 we show that they are *weighted* additions of local functions. Moreover, due to this two-step definition the desires can be redundant, because a desire does not add anything to the distinguished utility function when its constructible utility function ranks all worlds 0.

**Definition 7 (Nonmonotonic extension)** *A* (parameterized) desire *is defined by a pair of propositional formulas $a$ and $b$ together with the real $0 \leq p \leq 1$ for polarity, $l \in \{mM, MM, mm, Mm\}$ for lifting, and the real $s > 0$ for strength, and is denoted $D^{p}_{l:s}(a \mid b)$. A* (parameterized) desire specification *$DS = \{D^{p_1}_{l_1:s_1}(a_1 \,|\, b_1), \ldots, D^{p_n}_{l_n:s_n}(a_n \,|\, b_n)\}$ is a finite set of parameterized desires. The set of utility functions of $DS$, denoted by $U(DS)$, is the set of its models as given in Definition 1.*

$$U(DS) = \\ \{u \mid u \models D_{l_1:s_1}(a_1|b_1), \ldots, u \models D_{l_n:s_n}(a_n|b_n)\}$$

*The preferred or distinguished utility functions of a single desire, also called its local utility functions, are defined in two steps as follows. Let $u^{\alpha}_{a}$ be the utility function such that $u(w) = \alpha$ if $w \models a$, $u(w) = 0$ otherwise.*

$$CONS(D^{p}_{l:s}(a|b)) = \\ \{\{u^{\beta}_{a \wedge b} + u^{-\alpha}_{\neg a \wedge b} \mid \frac{\alpha}{\alpha + \beta} = p \text{ and } \alpha, \beta \geq 0\}$$

$$U_J(D^{p}_{l:s}(a|b)) = \\ U(\{D^{p}_{l:s}(a|b)\}) \cap CONS(D^{p}_{l:s}(a|b)) = \\ \{u^{\beta}_{a \wedge b} + u^{-\alpha}_{\neg a \wedge b} \mid \frac{\alpha}{\alpha + \beta} = p, \alpha, \beta \geq 0, \alpha + \beta \geq s\}$$

*The preferred or distinguished utility functions of a desire specification $DS$ are constructed as follows.*

$$CONS(DS) = \\ \left\{ u = u_1 + \ldots + u_n \;\middle|\; \begin{array}{l} u_1 \in CONS(D^{p_1}_{l_1:s_1}(a_1|b_1)), \\ \ldots, \\ u_n \in CONS(D^{p_n}_{l_n:s_n}(a_n|b_n)) \end{array} \right\}$$

$$U_J(DS) = U(DS) \cap CONS(DS)$$

The following proposition illustrates the formal construction by considering equivalent weighted additions, and it shows how to construct distinguished utility functions from single local utility functions instead of sets of them.

**Proposition 8 (Weighted additions)** *The constructible utility functions of*

$$DS = \{D^{p_1}_{l_1:s_1}(a_1|b_1), \ldots, D^{p_n}_{l_n:s_n}(a_n|b_n)\}$$

*are weighted additions of local utility functions.*

$$CONS(DS) = \\ \left\{ u = k_1 \times u_1 + \ldots + k_n \times u_n \;\middle|\; \begin{array}{l} u_1 \in U_J(D^{p_1}_{l_1:s_1}(a_1|b_1)), \\ \ldots, \\ u_n \in U_J(D^{p_n}_{l_n:s_n}(a_n|b_n)), \\ k_1 \geq 0, \ldots, k_n \geq 0 \end{array} \right\}$$

*The constructible utility functions of $DS$ are weighted additions of the minimal local utility functions $U_{min}(D^{p}_{l:s}(a|b)) = u^{s \times (1-p)}_{a \wedge b} + u^{-s \times p}_{\neg a \wedge b}$.*

$$CONS(DS) = \\ \left\{ u = k_1 \times u_1 + \ldots + k_n \times u_n \;\middle|\; \begin{array}{l} u_1 = U_{min}(D^{p_1}_{l_1:s_1}(a_1|b_1)) \\ \ldots \\ u_n = U_{min}(D^{p_n}_{l_n:s_n}(a_n|b_n)) \\ k_1 \geq 0, \ldots, k_n \geq 0 \end{array} \right\}$$

**Proof** *We first consider the first equivalence, and we prove that $CONS_1(DS) = CONS_2(DS)$ where $CONS_1$ is the construction defined in Definition 7 and $CONS_2$ is the first weighted addition defined above. That is, for each utility function in one construction we show for which variables $\alpha$, $\beta$ and $k$ this utility function is also part of the other construction.*

*⇒ For each desire, define $\alpha$, $\beta$ and $k_i$ in $CONS_2$ by $\alpha \times \frac{s}{\alpha+\beta}$, $\beta \times \frac{s}{\alpha+\beta}$ and $\frac{\alpha+\beta}{s}$ for $\alpha$ and $\beta$ in $CONS_1$. The local utility functions used in $CONS_2$ satisfy the constraints, because $\alpha \times \frac{s}{\alpha+\beta} + \beta \times \frac{s}{\alpha+\beta} = s$.*

*⇐ For each desire, define $\alpha$ and $\beta$ in $CONS_1$ by $k_i \times \alpha$ and $k_i \times \beta$ for $k_i$, $\alpha$ and $\beta$ in $CONS_2$.*

*We now consider the second equivalence, and we prove that $CONS_2(DS) = CONS_3(DS)$ where $CONS_2(DS)$ is again the first weighted addition defined above and $CONS_3$ is the second one. This follows directly from the fact that the utility function we constructed in the previous item is in fact the minimal one.*

*⇒ The ⇐-part of the previous item shows how to construct an element of $CONS_1$ from an element of $CONS_3$, and the ⇒-part of the previous item shows how to construct an element of $CONS_3$ from an element of $CONS_1$.*

*⇐ Trivial since $U_{min}$ is an element of $U_J$.*

The following proposition shows that the existence of distinguished utility functions of a desire specification does not follow from the existence of utility functions. Weydert has proven this for his defaults, i.e. for loss $D_{MM}$ desires. It is an open problem whether it can be proven in a more general context, e.g. for all $D_{MM}$ desires. This property is considered very desirable in reasoning about defaults (see [Kraus *et al.*, 1990]), but it is not clear whether it plays a similar role in reasoning about desires.

**Proposition 9** *Let DS be a desire specification. $U(DS) \neq \emptyset$ does* not *imply $U_J(DS) \neq \emptyset$.*

**Proof** *Two counterexamples are the desire specification $DS = \{D_{Mm}(p), D_{Mm}(\neg p)\}$ and the desire specification $DS = \{D_{MM}(p), D_{mm}(\neg p)\}$. Both have models but no preferred models.*

## 3 Conclusions

In this paper we have studied and extended the logic of desires in Lang's framework for qualitative decision theory. We introduced three parameters for the utilitarian desires that reflect its strength and the risk attitude of the agent, because utilities represent besides values also the agent's risk attitude. The parameterized desires can deal with the class of intuitions about the logical properties of desires by changing the parameter values for the requirements at hand. Despite the fact that the mechanisms developed in reasoning about defaults could be used for desires, it seems very unlikely that our logic of desires can be used to formalize defaults. In reasoning about uncertainty there is no formal counterpart of risk.

Subjects for further research are studies of minimization principles introduced in [Weydert, 1995; 1996; 1998] in the logic for desires, of existence theorems for fragments of the logic, and the search for general guidelines or heuristics for the values of the parameters (such as particular combinations of them) and for the determination of the parameter values in an interactive system.

## References

[Bacchus and Grove, 1996] F. Bacchus and A.J. Grove. Utility independence in a qualitative decision theory. In *Proceedings of KR'96*, pages 542–552, 1996.

[Boutilier, 1994] C. Boutilier. Toward a logic for qualitative decision theory. In *Proceedings of the KR'94*, pages 75–86, 1994.

[Doyle and Wellman, 1991] J. Doyle and M.P. Wellman. Preferential semantics for goals. In *Proceedings of AAAI-91*, pages 698–703, Anaheim, 1991.

[Haddawy and Hanks, 1992] P. Haddawy and S. Hanks. Representations for decision-theoretic planning: Utility functions for dead-line goals. In *Proceedings of the KR'92*, Cambridge, MA, 1992.

[Jeffrey, 1983] R. Jeffrey. *The Logic of Decision.* University of Chicago Press, 2nd edition, 1983.

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

[Kraus *et al.*, 1990] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Lang, 1996] J. Lang. Conditional desires and utilities - an alternative approach to qualitative decision theory. In *Proceedings of the ECAI'96*, pages 318–322, 1996.

[Luce and Raiffa, 1957] R.D. Luce and H. Raiffa. *Games and Decisions.* John Wiley, New York, 1957.

[Mullen, 1979] J.D. Mullen. Does the logic of preference rest on a mistake? *Metaphilosophy*, 10:247–255, 1979.

[Pearl, 1993] J. Pearl. From conditional ought to qualitative decision theory. In *Proceedings of the UAI'93*, pages 12–20, 1993.

[Tan and Pearl, 1994a] S.-W. Tan and J. Pearl. Qualitative decision theory. In *Proceedings of the AAAI'94*, 1994.

[Tan and Pearl, 1994b] S.-W. Tan and J. Pearl. Specification and evaluation of preferences under uncertainty. In *Proceedings of the KR'94*, pages 530–539, 1994.

[Thomason and Horty, 1996] R. Thomason and R. Horty. Nondeterministic action and dominance: foundations for planning and qualitative decision. In *Proceedings of the TARK'96*, pages 229–250. Morgan Kaufmann, 1996.

[van der Torre and Tan, 1999] L.W.N. van der Torre and Y.H. Tan. Contrary-to-duty reasoning with preference-based dyadic obligations. *Submitted*, 1999.

[Wellman and Doyle, 1992] M.P. Wellman and J. Doyle. Modular utility representation for decision-theoretic planning. In *Proceedings of the first international conference on artificial intelligence planning systems (AIPS92)*, pages 236–242, 1992.

[Weydert, 1995] E. Weydert. Default entailment a preferential construction semantics for defeasible inference. In Ipke Wachsmuth, Claus-Rainer Rollinger, and Wilfried Brauer, editors, *Annual German Conference on Artificial Intelligence (KI-19) : Bielefeld, Germany, September 11-13, 1995; proceedings*, volume LNAI 981, pages 173–184, Berlin, 1995. Springer.

[Weydert, 1996] E. Weydert. System J – revision entailment: Default reasoning through ranking measure updates. In Dov Gabbay and Hans Jürgen Ohlbach, editors, *Practical Reasoning - International Conference on Formal and Applied Practical Reasoning, FAPR'96*, volume 1085 of *Lecture Notes in Computer Science*, pages 637–649, Bonn, Germany, June 1996. Springer.

[Weydert, 1998] E. Weydert. System jz: How to build a canonical ranking model of a default knowledge base. In *Proceedings of the Seventh International conference on Knowledge Representation and Reasoning (KR'98)*, pages 190–201, 1998.