Applied Intelligence 14, 285–301, 2001

Parameters for Utilitarian Desires in a Qualitative Decision Theory

LEENDERT VAN DER TORRE
Department of Artificial Intelligence, Vrije Universiteit, Amsterdam, The Netherlands
torre@cs.vu.nl

EMIL WEYDERT
Max Planck Institute for Computer Science, 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. This paper is inspired from Lang's framework of qualitative decision theory, in which utility functions are constructed from desires. Unfortunately, there is no consensus about the desirable 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. We introduce three parameters, which help us to implement different facets of risk. The strength parameter encodes the importance of the desire, the lifting parameter encodes how to determine the utility of a set (proposition) from the utilities of its elements (worlds), 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.

Keywords: qualitative decision theory, QDT, agent theory, non-monotonic reasoning, BDI, desires

1. Introduction

Classical decision theory [1–3] has been developed to describe and prescribe rational human decision making. However, it seems that humans are frequently 'irrational'. For instance, they may consider the most normal situation together with (potentially) exceptionally highly preferred and dispreferred situations, and nothing in between (maybe a good model for resource bounded robots). Whereas this strategy may still be justified by the need to use computational resources economically, a rational reconstruction is not always possible, which complicates the description task. Here, we are mainly interested in the prescriptive role of decision theory and want to exploit it for decision making by artificial agents. For example, in decision-theoretic planning a robot receives requirements or imperatives, as well as knowledge, tries to figure out the corresponding set of admissible utility functions and probability distributions, calculates the expected utilities and acts accordingly. However, for this application domain of decision theory, a new problem arises. 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 should therefore be seen as well as *heuristic approximations* [4] as ways to *compactly* communicate our preferences and beliefs [5] that only refer to *qualitative abstractions* of utility functions and probability distributions (also called ranking distributions). 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.

Table 1. Requirements in decision-theoretic planning.

Utilities		Probabilities	
Quantitative Preference	Qualitative Desire	Quantitative Belief	Qualitative Default

In this paper we propose a logic of utilitarian desires that builds on previous work of Boutilier [6] and Lang [7]. 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 (help to) 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 [8–10]. In other words, which desires should be derived intuitively is unclear and may depend 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 choice 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 (proposition) from the utilities of its elements (worlds),
The polarity parameter encodes the proportion between gain of utility for rewards and loss of utility for violation.

Decision theory helps to elucidate the different intuitions about utilitarian desires and to justify and interpret 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 what may be called risk, whereas probabilities only encode frequencies.[1] Agents 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 also reflected by the contrived status of utility functions. With the parameters the risk component of each desire can be made more explicit,—we therefore call them risk parameters. The parametrization we propose for desires is not appropriate for defaults. Although Boutilier's and Lang's logics are analogous to formalisms proposed for defaults, as we show in greater detail for Lang's and Weydert's default reasoning approaches, they have as such been criticized in particular by [10, 11]. 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.

This paper is organized as follows. In Section 2 we repeat Lang's framework [7] and we compare it with Weydert's account of epistemic constructibility. In Section 3 we introduce the three parameters for desires, and in Section 4 we discuss different nonmonotonic constructions. In Section 5 we illustrate how the parameterized desires can formalize a large set of benchmark examples, some of which could not be formalized in the original framework.

2. Lang's Framework

Lang's framework for qualitative decision theory is based on an explicit construction of the agent's preference relation induced by a problem specification (desires + knowledge) via the definition of a class of utility function satisfying the desires. In this paper we only look at the desires and do not consider the representation of knowledge. The monotonic logic is based on Boutilier's modal framework for qualitative decision theory [6]. A desire for a if b, written as $D(a|b)$, is true if the best b worlds are a worlds. With utility functions these desires are defined as follows.

Definition 1 (Logic of desires [7]). Let $\mathcal{L}$ be a propositional language generated from a finite number of propositional variables, W the set of classical worlds associated with $\mathcal{L}$, and $Mod(a)$ the set of worlds satisfying

$a \in \mathcal{L}$. A conditional desire is defined by a pair of propositional formulas a and b and is denoted by $D(a|b)$. A desire specification is a finite set of conditional desires

$$DS = \{D(a_1 \mid b_1), \ldots, D(a_n \mid b_n)\}$$

An utility function u is a map from W to the reals $\mathbb{R}$. u satisfies the desire specification DS, written as $u \models DS$ (u is a model of DS), if and only if for $i = 1, \ldots, n$, the worlds in $Mod(b_i)$ maximizing u are worlds of $Mod(a_i)$.

Moreover, for each desire, a set of so-called local (distinguished) utility functions is introduced, and the (global) distinguished utility models of a desire specification are defined as the sum of local ones. The local utility functions of a desire represent a loss of utility (which is kept variable and denoted by α) associated with its violation, the violation of $D(a \mid b)$ being the negation of its materialization, that is $\neg a \wedge b$. The addition of local utility functions thus represents the addition of losses of utility, analogous to the addition of penalties in penalty logic [12]. The set of distinguished utility functions of a desire specification DS is a subset of all the utility models of DS, and thus gives rise to nonmonotonicity [13, 14]. This nonmonotonic extension is defined as follows.

Definition 2 (Nonmonotonic desires [7]). The real valued function $u_{a \mid b}$ is a local utility function of $D(a|b)$ if there exists an $\alpha > 0$ (its utility loss) such that

$$u_{a|b} = \begin{cases} 0 & \text{if } w \models b \to a \\ -\alpha & \text{if } w \models \neg a \wedge b \end{cases}$$

A distinguished utility function of DS is a real-valued function u such that:

1. $u \models DS$
2. $u = u_{a_1|b_1} + \cdots + u_{a_n|b_n}$

The distinguished utility functions can be interpreted as the most likely utility functions that satisfy the constraints 'the best b_i worlds are a_i worlds.' Moreover, the constraints can be interpreted as tools to impose restrictions on the choice process of the distinguished utility functions. The following two examples illustrate the kind of restrictions on the variables α that follow from these constraints. The first example of restriction on utility loss concerns a conflict which can be solved with the specificity principle.

Example 3 (Specificity). Consider the following desire specification [10], where $\top$ is a tautology.

$D(\neg s \mid \top)$	preference of no surgery over surgery
$D(s \mid i)$	inverse if surgery improves one's long term health

'Having surgery' (s) has utility loss α_1 and 'having no surgery if it improves one's long term health' ($\neg s \wedge i$) has utility loss α_2. The distinguished utility functions u are constructed as follows. The most specific desire has higher utility loss and thus overrides the more general one, because the restriction $\alpha_2 > \alpha_1$ follows from 'best i worlds are s worlds.'

$$u(w) = \begin{cases} 0 & \text{if } w \models \neg s \wedge \neg i \\ -\alpha_1 & \text{if } w \models s \\ -\alpha_2 & \text{if } w \models \neg s \wedge i \quad \text{with } \alpha_2 > \alpha_1 \end{cases}$$

Lang also shows that his nonmonotonic logic has the desirable property discussed in the defeasible reasoning community as 'inheritance to exceptional subclasses', i.e. it doesn't have the so-called 'drowning problem'. This means, given only two desires $D(a \mid b)$ and $D(a' \mid b)$ (with independent a, a'), the violation of a single one is better than the violation of both, which sounds reasonable.

The second example of restriction on utility loss, not discussed by Lang, is not concerned with specificity or conflicts (cf. Def. 10). For reasons that become apparent later on, we call it transitivity. To simplify, we exploit the possibility to restrict the set of all worlds to those which satisfy a set of formulas called the background knowledge—see [7] for details.

Example 4 (Transitivity). Consider the following desire specification together with the background knowledge $\neg(p \wedge c)$, $\neg(p \wedge h)$, $\neg(c \wedge h)$ and $(p \vee c \vee h)$. It tells us 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.

$D(p \mid p \vee c)$	I prefer to go to a party if I go to a party or to the cinema
$D(c \mid c \vee h)$	I prefer to go to the cinema if I go to the cinema or stay home

'Going to the cinema' (c) has utility loss α_1 and 'staying home' (h) has utility loss α_2, which leads to the following distinguished utility functions u. Note that the restriction $\alpha_2 > \alpha_1$ follows from 'best $c \vee h$ worlds are c worlds.'

$$\begin{aligned} u(w) = 0 \quad & \text{iff } w \models p \\ -\alpha_1 \quad & \text{iff } w \models c \\ -\alpha_2 \quad & \text{iff } w \models h \quad \text{with } \alpha_2 > \alpha_1 \end{aligned}$$

Finally, our next definition shows how the utility functions in Lang's system are a means to construct the agent's preference relation. We just give this definition to understand the role of the utility functions in his framework; we do not further consider these orders in this paper.

Definition 5 ($\geq_{DS}$ [7]). Let *DS* be a desire specification. $w_1 \geq_{DS} w_2$ if and only if all distinguished utility models u of *DS* satisfy $u(w_1) \geq u(w_2)$.

Boutilier's nonmonotonic inference strategy [6] to reason with defaults *and* desires is System Z, i.e. normality maximization. This approach has several drawbacks, among them the inability to deal with implicit independence assumptions (cf. inheritance to exceptional subclasses). On the other hand, Lang's addition of local utility functions is closely related to Weydert's epistemic construction paradigm for default reasoning [15–17], which we are now going to sketch.

A *default* is a conditional relationship between a pair of propositional formulas a and b, written as $b \Rightarrow a$, and expressing that 'b normally implies a'. A *default knowledge base DKB* is a finite set of defaults $\{b_1 \Rightarrow a_1, \ldots, b_n \Rightarrow a_n\}$. Ranking measures [18], in particular $\kappa\pi$-ranking measures, provide a natural and powerful semantic framework for default conditionals. They represent real-valued measures of surprise, or implausibility rankings, which generalize Spohn's integer-valued κ-rankings [19] and the [0, 1]-valued possibility measures of possibilistic logic [20]. Formally speaking, a *$\kappa\pi$-ranking measure* is a function from the set of propositions (world sets) 2^W (W finite) to the positive reals $\mathbb{R}^+$ and ∞ which satisfies $R(W) = 0$, $R(\emptyset) = \infty$, and $R(a \cup b) = \min\{R(a), R(b)\}$.[2] The uniform ranking R_0 assigns 0 to each world. The higher the value of a proposition, the more exceptional or implausible it is. There are basically two ways to define a satisfaction relation $\models$ between $\kappa\pi$-rankings R and defaults $b \Rightarrow a$ (we assume $\max(\emptyset) = \infty$).

1. $R \models_{\geq 1} b \Rightarrow a$ iff $\max_{w \models a \wedge b} R(w) + 1 \leq \max_{w \models \neg a \wedge b} R(w)$ (weak).
2. $R \models_{>0} b \Rightarrow a$ iff $\max_{w \models a \wedge b} R(w) < \max_{w \models \neg a \wedge b} R(w)$ (strong).

These definitions are equivalent for integer-valued measures, but differ for real-valued ones. In particular, they do not support the same conditional axioms. The first truth condition may be preferable because it offers more flexibility and is technically easier to handle. The second truth condition, however, is closer to the classical maximal world semantics for conditionals. We come back to this point below.

Default reasoning in the context of the $\kappa\pi$-ranking-semantics is usually based on a preferred model concept. In Weydert's basic framework, a proposition a is nonmonotonically entailed by a proposition b and a default knowledge base *DKB* if and only if for every *preferred* $\kappa\pi$-ranking R (weakly) satisfying *DKB* ($R \models_{\geq 1}$), R (strongly) satisfies $b \Rightarrow a$ ($R \models_{>0} b \Rightarrow a$). For example, system Z or Z^+ pick up a single preferred model, namely the uniquely determined pointwise minimal $\kappa\pi$-ranking weakly satisfying *DKB*.

The idea behind Weydert's preferred construction paradigm has been to prefer those $\kappa\pi$-ranking models of *DKB* which can be obtained from the uniform $\kappa\pi$-ranking R_0 by iterated Spohn revision [19], i.e. Jeffrey-conditionalization, with material implications $b \rightarrow a$ corresponding to defaults $b \Rightarrow a \in DKB$. If $R(b \rightarrow a) = 0$, revising R by $b \rightarrow a$ can be achieved by 'shifting' the abnormality part $b \wedge \neg a$ upwards, i.e. by making the $b \wedge \neg a$-worlds uniformly more exceptional. The resulting $\kappa\pi$-ranking is denoted by $R[b \wedge \neg a + \alpha]$, where $\alpha \geq 0$. $R[b \wedge \neg a + \alpha]$ is the unique $\kappa\pi$-ranking R' such that $R'(w) = R(w) + \alpha$ for $w \models b \wedge \neg a$, and $R'(w) = R(w)$ otherwise.

Definition 6 (Constructibility [15–17]). Let $DKB = \{b_1 \Rightarrow a_1, \ldots, b_n \Rightarrow a_n\}$ be a set of defaults. A $\kappa\pi$-ranking R is called epistemically constructible over *DKB* if and only if

- $\exists \alpha_1, \ldots, \alpha_n \geq 0$:

$$R = R_0[b_1 \wedge \neg a_1 + \alpha_1] \cdots [b_n \wedge \neg a_n + \alpha_n].$$

It is called strictly epistemically constructible over *DKB* if and only if all the α_i can be chosen $\alpha_i > 0$.

Weydert has also discussed stronger notions of constructibility, up to the canonical JZ-construction in [17]. This has resulted in a hierarchy of increasingly strong, well-behaved default inference concepts. Here, we are mainly concerned with the most cautious variant, called

J-entailment,[3] which picks up as preferred all the epistemically constructible $\kappa\pi$-models of *DKB*.

The following Proposition 7 shows how the distinguished utility models of desires in Definition 1 correspond to the preferred constructible $\kappa\pi$-models according to Definition 6.

Proposition 7. *Let $DS = \{D(a_i \mid b_i) \mid 1 \leq i \leq n\}$ be a desire specification, i.e. a finite set of conditional desires, and let $DKB = \{b_i \Rightarrow a_i \mid 1 \leq i \leq n\}$ be the corresponding default knowledge base. Also let DS and DKB stay consistent if we disallow the ∞-value.*

- *If a real-valued utility function u is a distinguished utility function of DS, then $R = -u$ defines a constructible $\kappa\pi$-ranking model of DKB; the converse only holds if R is strictly constructible.*
- *If for all constructible $\kappa\pi$-ranking models R of DKB we have $R \models_{>0} b \Rightarrow a$ then we have for all distinguished utility functions u that $u \models D(a \mid b)$; the converse only holds if R is strictly constructible.*

Proof:

$\Rightarrow$:

We only prove the first item, because it directly entails the second one. For the first point, the crucial fact is that the shifting order, the order of the updates in Definition 6 is irrelevant. The result now follows from the following two observations.

1. An $[a + \alpha]$-update corresponds exactly to the addition of a local utility function with utility loss α for a.
2. The initial ranking R_0 corresponds to the assumption that nothing more than the local utility functions contributes to the global utility function on W.

$\Leftarrow$:

The converse holds if we assume strict constructibility. Because the requirement $\alpha_i > 0$ guarantees a correspondence between $[a+\alpha]$-updates and the addition of a local utility function with utility loss α for a. On the other hand, it fails if updates can be redundant in the sense that the α_i's are allowed to be zero, whereas utility loss is always strictly positive. We give a counterexample to the converse of the second item, which also disproves the converse of the first one. Consider the desire specification $DS = \{D(p \wedge q \mid \top),\ D(p \mid \top)\}$ and the corresponding default knowledge base $DKB = \{\top \Rightarrow p \wedge q,\ \top \Rightarrow p\}$. For all distinguished utility models u we have $u \models D(p \mid \neg(p \wedge q))$, but we do not have for all constructible $\kappa\pi$-ranking R of DKB that $R \models \neg(p \wedge q) \Rightarrow p$. A counterexample of the latter is the ranking $R = R_0[\neg(p \wedge q) + 1][\neg p + 0]$. □

The previous proposition shows that the two formal systems are rather similar, the main difference being the notion of redundancy.[4] On the other hand, the underlying intuitions are very different: the $\kappa\pi$-ranking are order-of-magnitude abstractions, whereas the utility functions are not. The more sophisticated extensions proposed by Weydert can also be used in Lang's framework to further restrict the set of distinguished utility functions. But first we introduce the three parameters.

3. Risk Parameters

In this section we show how desires can be parameterized. We introduce three parameters. The first one encodes explicit strength and the other two reflect the attitude towards risk. With the risk parameters we can discriminate between optimistic agents that assume that the best possible state will arise and reason about gain of utility or rewards, pessimistic agents that assume that the worst state will arise and reason about loss of utility or penalties, as well as different intermediate attitudes. In this sense the risk parameters are related to the distinction between monopolarity and bipolarity as discussed in [6, 11]. We first discuss the monotonic logic before we turn in the next section to the nonmonotonic construction.

3.1. The Logic: Explicit Strengths

In this section we introduce the first parameter, which represents the strength s of a desire. Because utility functions are real-valued, we allow s to be an arbitrary strictly positive real. The truth conditions for parameterized desires are similar to those offered by Weydert (or system Z^+, using integers) for default conditionals. But we follow Lang by assigning negative values to violations.

$$u \models D_{\geq s}(a \mid b)$$
$$\text{iff } \max_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w)$$

In Example 25 we show how desires with a higher strength can override desires with a lower strength. In particular, we show that under certain circumstances, from the strong desire 'to be healthy' and the weaker desire 'to be wealthy' we can infer that being healthy and poor is preferred to being unhealthy and wealthy.

3.2. 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. [21]), 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.[5] Let us call these operators $Mu(p)$ and $mu(p)$. If there are no worlds satisfying p, then the maximum and the minimum are respectively $-\infty$ and ∞.

$$Mu(p) = \max_{w \models p} u(w)$$
$$mu(p) = \min_{w \models p} u(w)$$

$Mu(p)$ 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 $Mu(p) \geq u(w) \geq mu(p)$. If we know that we are in a p state, then assuming $Mu(p)$ is optimistic (the best case arises, maybe we can enforce it deterministically by our actions) and assuming $mu(p)$ is pessimistic (the worst case arises, maybe it is enforced by our adversary).

$Mu(p)$ and $mu(p)$ can be used to define different types of constraints for desires (with strength s). In fact, these two poles can be compared in the following four ways, assuming that a_1 and a_2 are nonempty propositions.

$$u \models a_1 \succ_{mM:s} a_2$$
$$\Leftrightarrow mu(a_1) \geq s + Mu(a_2)$$
$$\Leftrightarrow \min_{w \models a_1} u(w) \geq s + \max_{w \models \neg a_2} u(w)$$

$$u \models a_1 \succ_{MM:s} a_2$$
$$\Leftrightarrow Mu(a_1) \geq s + Mu(a_2)$$
$$\Leftrightarrow \max_{w \models a_1} u(w) \geq s + \max_{w \models \neg a_2} u(w)$$

$$u \models a_1 \succ_{mm:s} a_2$$
$$\Leftrightarrow mu(a_1) \geq s + mu(a_2)$$
$$\Leftrightarrow \min_{w \models a_1} u(w) \geq s + \min_{w \models \neg a_2} u(w)$$

$$u \models a_1 \succ_{Mm:s} a_2$$
$$\Leftrightarrow Mu(a_1) \geq s + mu(a_2)$$
$$\Leftrightarrow \max_{w \models a_1} u(w) \geq s + \min_{w \models \neg a_2} u(w)$$

In Definition 8 below a desire $D(a \mid b)$ is defined as usual by $a \wedge b \succ \neg a \wedge b$ (following von Wright's expansion rule). If b is inconsistent, i.e. if there are no worlds satisfying it, then we assume that the desire is vacuously true.

Definition 8 (Logic of parameterized desires). A (parameterized) 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 the lifting strategy, and is denoted $D_{l:s}(a \mid b)$. A (parameterized) desire specification $DS = \{D_{l_1:s_1}(a_1 \mid b_1), \ldots, D_{l_n:s_n}(a_n \mid b_n)\}$ is a finite set of parameterized desires. An utility function u, i.e. a map from W to the reals $\mathbb{R}$, satisfies the desire $D_{l:s}(a \mid b)$, written as $u \models D_{l:s}(a \mid b)$, if and only if there are no $a \wedge b$ worlds, or there are no b worlds, or the truth conditions corresponding to $l : s$ hold.

$$u \models D_{mM:s}(a \mid b)$$
$$\Leftrightarrow mu(a \wedge b) \geq s + Mu(\neg a \wedge b)$$
$$\Leftrightarrow \min_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w)$$

$$u \models D_{MM:s}(a \mid b)$$
$$\Leftrightarrow Mu(a \wedge b) \geq s + Mu(\neg a \wedge b)$$
$$\Leftrightarrow \max_{w \models a \wedge b} u(w) \geq s + \max_{w \models \neg a \wedge b} u(w)$$

$$u \models D_{mm:s}(a \mid b)$$
$$\Leftrightarrow mu(a \wedge b) \geq s + mu(\neg a \wedge b)$$
$$\Leftrightarrow \min_{w \models a \wedge b} u(w) \geq s + \min_{w \models \neg a \wedge b} u(w)$$

$$u \models D_{Mm:s}(a \mid b)$$
$$\Leftrightarrow Mu(a \wedge b) \geq s + mu(\neg a \wedge b)$$
$$\Leftrightarrow \max_{w \models a \wedge b} u(w) \geq s + \min_{w \models \neg a \wedge b} u(w)$$

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 $u(w_1) > u(w_2)$.

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

Note that the semantics of MM-desires can be traced back to Lewis [22]. The following proposition shows the relations between the different types of desires.

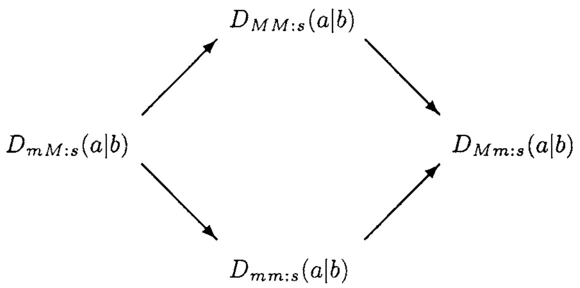

Figure 1. Relations between the four types of desires.

Proposition 9 (Relations between desires). *We have the following relations between the parameterized desires based on different values for the lifting parameter.*

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

These relations are represented in Fig. 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 $Mu(a) \geq mu(a)$. □

Suppose that the lifting parameters of all desires in 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 can be realized by indexing–instead of the individual desires–the satisfiability relation with the used lifting parameter, e.g. $\models_{mM}$. A desire specification *DS* is said to be an *l*-conflict set if *DS* is inconsistent with respect to $\models_l$.

Definition 10 (Conflicts). A desire specification *DS* (with only strength parameters) is an *l*-conflict set if there is no *u* with $u \models D_{l:s}(a \mid b)$ for each $D_s(a \mid b) \in DS$. *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 $D_{mm:s}(a \mid b)$ is the dual of $D_{MM:s}(a \mid b)$ and has similar properties. As we already observed above, $D_{mm:s}(a \mid b)$ reflects a pessimistic view in the sense that it only considers the worst *b* states, whereas $D_{MM:s}(a \mid b)$ only considers the best *b* states.

Secondly, the desire $D_{mM:s}(a \mid b)$ 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, the two desires 'to be healthy' $D_{s_1}(h \mid \top)$ and 'to be wealthy' $D_{s_2}(w \mid \top)$ always constitute 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 second desire prefers the first world to the second one and the second desire goes the other way around. Moreover, the set of surgery desires in Example 3 is an *mM*-conflict set.

However, there are also situations where strong desires can be useful. A non-trivial conflict-free instance is the transitivity set discussed in Example 4. We reconsider it below, together with two other conflict-free sets of desires.

Example 11 (Transitivity, continued). Consider the following three desire specifications, with the background knowledge that *p*, *c* and *h* are mutually exclusive and exhaustive. We also give the representation based on $\succ$ operators, because they are the most readable. *CTD* and *ATD* represent contrary-to-duty and according-to-duty examples extensively discussed in the logic of obligations, see e.g. [23, 24].

TRANS

$D_{mM:1}(p \mid p \vee c)$	$p \succ_{mM:1} c$
	$mu(p) \geq 1 + Mu(c)$
$D_{mM:1}(c \mid c \vee h)$	$c \succ_{mM:1} h$
	$mu(c) \geq 1 + Mu(h)$

CTD

$D_{mM:1}(p \mid p \vee c \vee h)$	$p \succ_{mM:1} c \vee h$
	$mu(p) \geq 1 + Mu(c \vee h)$
$D_{mM:1}(c \mid c \vee h)$	$c \succ_{mM:1} h$
	$mu(c) \geq 1 + Mu(h)$

ATD

$D_{mM:1}(p \mid p \vee c)$	$p \succ_{mM:1} c$
	$mu(p) \geq 1 + Mu(c)$
$D_{mM:1}(p \vee c \mid p \vee c \vee h)$	$p \vee c \succ_{mM:1} h$
	$mu(p \vee c) \geq 1 + Mu(h)$

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 $u(w_1) \geq 1 + u(w_2) \geq 2 + u(w_3)$.

Moreover, this strongest desire concept may be useful in combination with further restrictions on the sets of worlds to be compared. For example, we can restrict the constraint to the most normal $a \wedge b$ and $\neg a \wedge b$ worlds. In the definition below the models are assumed to contain, besides an utility function, also a plausibility or $\kappa\pi$-ranking to determine the most normal worlds. We write $N(a)$ for the most normal worlds satisfying a, i.e. $N(a) = \{w \mid p(w) = \max_{w' \models a} p(w')$.

$$(u, p) \models D_{mM:s} N(a \mid b)$$
$$\Leftrightarrow mu_n(a \wedge b) \geq s + Mu_n(\neg a \wedge b)$$
$$\Leftrightarrow \min_{w \in N(a \wedge b)} u(w) + s \geq s + \max_{w \in N(\neg a \wedge b)} u(w)$$

The following example is adapted from the cottage housing regulations investigated in the logic of obligations [24] and illustrates the new desires. See [7] for a discussion on the relation between the logic of desires and the logic of obligations.

Example 12 (Fence and dog). Consider the following desire specification.

$D_{mM:s_1} N(\neg f \mid \top)$	preference for no fence
$D_{mM:s_2} N(f \mid d)$	preference for fence if there is a dog
$D_{mM:s_3} N(d \mid \top)$	preference for a dog

The specification is consistent, though

$$\{D_{s_1}(\neg f \mid \top), D_{s_2}(f \mid d), D_{s_3}(d \mid \top)\}$$

is an mM-conflict set. The first two desires $D_{mM:s_1} N(\neg f \mid \top)$ and $D_{mM:s_2} N(f \mid d)$ imply that either 'having a dog and no fence' ($d \wedge \neg f$) or 'having a dog and a fence' ($d \wedge f$) is exceptional, because if there would be most normal $d \wedge \neg f$ and $d \wedge f$ worlds then according to the first desire the first worlds would be preferred, and according to the second desire the second worlds.

A second way to adapt the strongest desire $D_{mM}(a \mid b)$ is to restrict the set of compared worlds with a ceteris paribus constraint, see e.g. [4, 11, 25, 26] for details. For example, consider a set of circumstances $\{c_1, \ldots, c_n\}$ where each c_i formalizes similar circumstances except for the value of a. We can introduce a ceteris paribus constraint through

$$u \models D_{l:s} CP(a \mid b)$$
$$\Leftrightarrow u \models D_{l:s}(a \mid b \wedge c_1) \wedge \cdots \wedge D_{l:s}(a \mid b \wedge c_n)$$

The major problem with ceteris paribus preferences, see e.g. [10, 11], is that they cannot formalize the common sense intuition in Example 3 that more specific preferences override more general conflicting ones. Tan and Pearl [26] propose a way to deal with specificity in an extension of ceteris paribus preferences with priorities, but Bacchus and Grove [10] criticize that they are 'incorporating specificity without changing the underlying semantics.' A further discussion of these two extensions based on most normal worlds and ceteris paribus preferences is outside the scope of this paper.

Finally, we consider the weakest desire $D_{Mm:s}(a \mid b)$. 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, we show later in this paper (in Example 28 and 29) how a desire based on such a weak constraint in combination with an appropriate nonmonotonic construction may have a considerable impact. On the other hand, some examples also suggest that the three other notions may be too strong. This can be illustrated by the marriage of Sue example of Bacchus and Grove [10].

Example 13 (Marriage). Consider the desire specification *DS* that consists of the following three desires.

$D_1(j \mid \top)$	Sue prefers to be married to John over not being married to him
$D_1(f \mid \top)$	Sue prefers to be married to Fred over not being married to him
$D_1(\neg(j \wedge f) \mid \top)$	Sue prefers not to be married to both

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

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

is inconsistent, because 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 $D_{Mm:1}(j \mid \top)$, $D_{Mm:1}(f \mid \top)$ and $D_{Mm:1}(\neg(j \wedge f) \mid \top)$ is

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

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 desire specification of Example 12.

Example 14 (Fence and dog). Consider the desire specification *DS* that consists of the following three desires.

$D_1(\neg f \mid \top)$	preference for no fence
$D_1(f \mid d)$	preference for fence if there is a dog
$D_1(d \mid \top)$	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 $D_{Mm:1}(\neg f \mid \top)$, $D_{Mm:1}(f \mid d)$ and $D_{Mm:1}(d \mid \top)$ is

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

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 $u \models D_{Mm:1}(f \mid d)$ because $f \wedge d$ worlds are better than $\neg f \wedge d$ worlds, and we have $u \models D_{Mm:1}(d \mid \top)$ 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 desire notion, and there are desire specifications which can only be analyzed with the weakest desire concepts. However, examples without inherent conflicts usually can be more naturally formalized with D_{MM}, i.e. with the semantics used in Boutilier's, Lang's and Weydert's frameworks. This optimistic assumption can be justified, for example, by the implicit assumption that all worlds are feasible in the sense that the agent has the power to reach any of the physically possible worlds [7]. This will therefore be our standard representation.

The fence and dog example also illustrates that sometimes there are several ways to represent a set of desires, for example as strong constraints restricted to the most normal worlds, or as weak constraints. Likewise, the marriage example can be represented as a strong constraint restricted to the most normal worlds (if the worlds in which she is married to neither are the only most normal worlds). Which representation is the best one depends, just like the value of our parameters, on the situation being formalized.

3.3. *The Logic: Bipolarity*

In this section we introduce our third parameter p, which expresses the relation between gain of utility for rewards and loss of utility for violations. We call it the polarity parameter. Desires with polarity are denoted by $D^p_{l:s}(a \mid b)$. The polarity parameter is used for local utility functions, i.e. in the construction of the distinguished utility functions. We may now consider local utility functions that take into account not only 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 or model of $D_{ls}(a \mid b) - u_{a|b}$ 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

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

For representational convenience we represent this utility function by $u = u^{\beta}_{a\wedge b} + u^{-\alpha}_{\neg a \wedge b}$. The polarity parameter is defined by $p = \frac{\alpha}{\alpha+\beta}$, and thus restricts the relative values of α and β. Obviously we have $0 \leq p \leq 1$. For example, mixed gain-loss desires with polarity 0.5 have local utility functions u of the form $u = u^{\alpha}_{a\wedge b} + u^{-\alpha}_{\neg a\wedge b}$, with $\alpha \geq 0.5 \times s$, i.e.

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

Note that the polarity parameter only affects the local utility functions, and that consequently two desires that only differ in their polarity have the same monotonic semantics. In the next section we show that they only differ in their nonmonotonic semantics.

If the polarity of a desire is 0 then we call it a gain desire, because its utility loss α is zero. Likewise, if its polarity is 1 then we call it a loss desire, because its utility gain β is zero. Finally, in contrast to Lang's original proposal we allow α and β to be 0. In the following section we show how the parameters influence the construction of preferred utility models.

4. The Nonmonotonic Extension

The philosophy of Lang's framework is to define the global distinguished utility functions satisfying a set of desires as a function of the local (distinguished) utility functions of its elements. A similar philosophy underlies multi-attribute utility theory with the use of additive independence [3, 10, 27]. There are several different ways to realize the idea of constructing preferred utility models of a desire specification from (preferred) utility models of its elements. In this paper we reformulate and generalize Lang's approach in a model preference semantics similar to the one adopted in Weydert's preferential construction paradigm for default reasoning. It is slightly more flexible, and allows us to take a broader, more integrated perspective.

4.1. Constructions

Our general strategy for defining the preferred utility models of a desire specification *DS* is a very simple one. First, we collect all the utility functions which are suitably constructible over *DS*, then we take the intersection with the set of all *DS* utility models $U(DS)$. The most basic procedure, close to Lang's perspective, is to consider the set of all the (weakly) constructible utility functions over *DS*, which is represented by $CONS(DS)$. The corresponding set of preferred or distinguished utility models $U_J(DS)$ is now obtained by intersecting $U(DS)$ and $CONS(DS)$. *J* refers to Weydert's J-entailment, which got its name from Jeffrey-conditionalization. In this framework, the distinguished utility functions are *weighted* additions of local utility functions and these weights may vanish when the desires are redundant.

Definition 15 (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 $DS = \{D^{p_1}_{l_1:s_1}(a_1 \mid b_1), \ldots, D^{p_n}_{l_n:s_n}(a_n \mid b_n)\}$ is a finite set of parameterized desires. $u \models D^p_{l:s}(a \mid b)$ is set equivalent to $u \models D_{l:s}(a \mid b)$ (Definition 8). The set of utility models of *DS* is denoted by $U(DS)$, and given by

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

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

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

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

The distinguished utility models of a desire specification *DS* are constructed as follows.

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

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

In Section 5, we are going to illustrate the nonmonotonic extension by several benchmark examples. In the remainder of this section we show some properties of our new logic and in Section 4.2 we sketch an extension of the nonmonotonic construction based on ideas of minimal constructibility as discussed in [16, 17].

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

Proposition 16 (Weighted additions). *The constructible utility models of*

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

are weighted additions of local utility functions.

$$CONS_1(DS) = \left\{ \begin{array}{l} u = k_1 \times u_1 + \cdots + k_n \times u_n \\ \quad u_1 \in U_J(D^{p_1}_{l_1:s_1}(a_1|b_1)) \\ \quad \ldots \\ \quad u_n \in U_J(D^{p_n}_{l_n:s_n}(a_n \mid b_n)) \\ \quad k_1 \geq 0, \ldots, k_n \geq 0 \end{array} \right\}$$

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

$$CONS_2(DS) = \left\{ \begin{array}{l} u = k_1 \times u_1 + \cdots + k_n \times u_n \\ \left| \begin{array}{c} u_1 = U_{min}\big(D^{p_1}_{l_1:s_1}(a_1 \mid b_1)\big) \\ \cdots \\ u_n = U_{min}\big(D^{p_n}_{l_n:s_n}(a_n \mid b_n)\big) \\ k_1 \geq 0, \ldots, k_n \geq 0 \end{array} \right. \end{array} \right\}$$

Proof: We start with the first equivalence and prove that $CONS_1(DS) = CONS(DS)$, where $CONS$ is the construction from Definition 15. That is, for each utility function in one construction we show for which variables α, β and k_i this utility function is also part of the other construction.

($\Rightarrow$) For each desire, define α, β and k_i in $CONS_1$ by $\alpha \times \frac{s}{\alpha+\beta}$, $\beta \times \frac{s}{\alpha+\beta}$ and $\frac{\alpha+\beta}{s}$ for α and β in $CONS$. The local utility functions used in $CONS_1$ satisfy the constraints, because $\alpha \times \frac{s}{\alpha+\beta} + \beta \times \frac{s}{\alpha+\beta} = s$.

($\Leftarrow$) For each desire, define α and β in $CONS_1$ by $k_i \times \alpha$ and $k_i \times \beta$ for k_i, α and β in $CONS_1$.

We continue with the second equivalence and prove that $CONS_2(DS) = CONS_1(DS)$. This follows directly from the fact that the utility function we constructed in the previous item is in fact the minimal one.

($\Rightarrow$) The $\Leftarrow$-part of the previous item shows how to construct an element of $CONS$ from an element of $CONS_2$, and the $\Rightarrow$-part shows how to construct an element of $CONS_2$ from one of $CONS_1$.

($\Leftarrow$) Trivial since the U_{min}-model is an U_J-model. □

We now formalize redundancy with respect to a desire specification. A desire is redundant if adding this desire to the desire specification does not change the set of distinguished utility functions.

Definition 17 (Redundancy). A desire D is redundant for the desire specification DS if and only if $U_J(DS) = U_J(DS \cup \{D\})$.

It is obvious that duplicate desires are redundant.

Proposition 18 (Redundancy of duplicates). *For $D \in DS$, D is redundant for DS.*

The following example illustrates that redundancy is a much stronger notion than logical derivibility, and is not restricted to duplicates.

Example 19 (Redundancy). The desire $D^p_{l:s}(p \wedge q \mid \top)$ derives the desire $D^p_{l:s}(p \mid \top)$, but the latter is not redundant for the former. For example, the utility function $u^{-1}_{\neg p\vee\neg q} + u^{-1}_{\neg p}$ is not a distinguished utility model of $\{D^1_{MM:1}(p \wedge q \mid \top)\}$ but it is one of the desire specification $\{D^1_{MM:1}(p \wedge q \mid \top),\ D^1_{MM:1}(p \mid \top)\}$.

The desire for $D^1_{l:s}(a \mid \top)$ is redundant for the desire specification $DS = \{D^1_{l:s}(a \mid b),\ D^1_{l:s}(a \mid \neg b)\}$. Call the utility losses respectively α, α_1 and α_2. The distinguished utility models of DS are $u^{-\alpha_1}_{\neg a\wedge b} + u^{-\alpha_2}_{\neg a\wedge\neg b}$ and the distinguished utility models of $DS \cup \{D^1_{l:s}(a \mid \top)\}$ are $u^{-\alpha_1-\alpha}_{\neg a\wedge b} + u^{-\alpha_2-\alpha}_{\neg a\wedge\neg b}$. Define α_1 and α_2 in the former by $\alpha_1 + \alpha$ and $\alpha_2 + \alpha$ for α_1, α_2 and α in the latter.

We now consider the equivalence under translations or global shifting.

Definition 20 (Equivalence). Two sets of utility functions U_1 and U_2 are equivalent, written as $U_1 \sim U_2$, if and only if for each $u_2 \in U_2$ there is a $u_1 \in U_1$ and a uniform $u^{\alpha}_{\top}$ such that $u_1 = u_2 + u^{\alpha}_{\top}$, and vice versa.

A set of utility functions U is closed under global shifting if and only if $u \in U$ implies $u + u^{\alpha}_{\top} \in U$ for all $\alpha \in \mathbb{R}$.

The full utility models sets of desire specifications are closed under global shifting.

Proposition 21 (Closure under global shifting). *Let DS be a desire specification. $U(DS)$ is closed under global shifting.*

Proof: The desires only refer to relative utilities. □

The next proposition shows that a global shift of local utility functions gives equivalent distinguished utility functions, and that the distinguished utility functions of the closure of the constructible utility functions is exactly the closure of the distinguished utility functions of the desire specification. Consequently the 0 in the local utility functions of Definition 15 does not play a distinctive role.

Proposition 22 (Closure under global shifting, continued). *For some γ, let $CONS^{\gamma}(D^p_{l:s}(a \mid b))$ be the global shift of the constructible utility functions of a single desire, i.e. $CONS^{\gamma}(D^p_{l:s}(a \mid b))$ =*

$$\big\{u + u^{\gamma}_{\top} \,\big|\, u \in CONS\big(D^p_{l:s}(a \mid b)\big)\big\},$$

and let $U_J^{\gamma}(D_{l:s}^{p}(a|b))$, $CONS^{\gamma}(DS)$ *and* $U_J^{\gamma}(DS)$ *be defined as in Definition* 15, *replacing* $CONS(D_{l:s}^{p}(a \mid b))$ *by* $CONS^{\gamma}(D_{l:s}^{p}(a|b))$. *Then we have*:

- $U_J^{\gamma}(DS) \sim U_J(DS)$, *i.e. they are equivalent under global shifting.*

$$\begin{aligned} &\textit{Let } CONS^{*}\big(D_{l:s}^{p}(a|b)\big) \\ &\quad = \cup\big\{CONS^{\gamma}\big(D_{l:s}^{p}(a \mid b)\big) \mid \gamma \in \mathbb{R}\big\} \end{aligned}$$

be the closure under arbitrary global shifts of the constructible utility functions of a single desire, i.e. $CONS^{*}(D_{l:s}^{p}(a \mid b)) =$

$$\big\{u + u_{\top}^{\gamma} \mid u \in CONS\big(D_{l:s}^{p}(a \mid b)\big)\big\},$$

and let $U_J^{*}(D_{l:s}^{p}(a|b))$, $CONS^{*}(DS)$ *and* $U_J^{*}(DS)$ *be defined accordingly, replacing* $CONS(D_{l:s}^{p}(a \mid b))$ *by* $CONS^{*}(D_{l:s}^{p}(a|b))$.

- $U_J^{*}(DS)$ *is closed under global shifting, i.e. for any* α *if* $u \in U_J^{*}(DS)$ *then* $u + u_{\top}^{\alpha} \in U_J^{*}(DS)$.
- $U_J^{*}(DS)$ *is the closure under global shifting of* $U_J(DS)$.

Proof: Follows directly from the fact that $k \times (u_1 + u_2) = k \times u_1 + k \times u_2$. □

It is possible to show that the existence of distinguished utility models of a desire specification does not follow from the existence of utility models, i.e. consistency.[6] The model existence property (or consistency preservation) is considered very desirable in reasoning about defaults (see [14]), but it is not clear whether it plays a similar role in reasoning about desires.

Proposition 23. *There are desire specifications DS such that* $U(DS) \neq \emptyset$ *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. □

4.2. Justified Constructions

Finally we want to show a further strengthening of the logic of desires. The question is what principles can we accept to restrict or focus the set of preferred utility models even further. The usual approach in nonmonotonic reasoning is to select rankings that are most compact. For example, Tan and Pearl [11] use a global minimization principle. However, this has several drawbacks, see e.g. [28]. An alternative approach is based on Weydert's preferential construction paradigm. In particular, he has investigated the idea of minimal constructibility [16]. In this paper, we only consider a variant of so-called justified J-entailment or JJ-entailment. However, other more complex refinements, based on more sophisticated notions of minimality, can be used in our framework too. Justified constructions are most easily defined using Proposition 16, which realizes the constructible utility functions of *DS* as weighted additions of the minimal local utility functions.

$$CONS(DS) = \left\{ \begin{array}{l|c} & u = k_1 \times u_1 + \cdots + k_n \times u_n \\ & u_1 = U_{min}\big(D_{l_n:s_1}^{p_1}(a_1 \mid b_1)\big) \\ & \cdots \\ & u_n = U_{min}\big(D_{l_n:s_n}^{p_n}(a_n \mid b_n)\big) \\ & k_1 \geq 0, \ldots, k_n \geq 0 \end{array} \right\}$$

Definition 24. An element of *CONS*(*DS*) is justified with respect to *DS* if and only if for each $k_i > 0$ the corresponding desire constraint holds as an equality constraint, i.e. when replacing $\geq$ by $=$. That is, for $D_{l:s}^{p}(a \mid b)$, we have for $l = MM$:

$$l = MM : \max_{w \models a \wedge b} u(w) = s + \max_{w \models \neg a \wedge b} u(w)$$

We set $U_{JJ}(DS) = \{u \in U_J \mid u \text{ is justified w.r.t } DS\}$

In the following section we illustrate the justified distinguished utility functions by several examples.

5. Examples

In this section we illustrate how parameterized desires can formalize several benchmark examples of qualitative decision theory. We consider three different lifting policies: first we illustrate Lang's D_{MM}-desires, then we consider the stronger D_{mM}-desires, and finally we look at the weaker D_{Mm}-desires. Moreover, concerning polarity we focus on mixed gain-loss desires which are compared in several examples with the loss desires from Lang's framework. For convenience we use the second representation from Proposition 16 based on weighted additions and the construction of justified distinguished utility functions in Definition 24.

The following first example of D_{MM}-desires has been discussed when we introduced the strong desires D_{mM} in Section 3.2. It contains two simple desires that do not interact (in contrast to following examples) and illustrates the strength parameter.

Example 25 (Healthy and wealthy). Consider the following desires.

$$D^{0.5}_{MM:4}(h) \quad Mu(h) \geq 4 + Mu(\neg h)$$
$$D^{0.5}_{MM:2}(w) \quad Mu(w) \geq 2 + Mu(\neg w)$$

The distinguished and justified distinguished utility functions are constructed as follows. The latter are written between the brackets '(' and ')'. In our examples, there is always a uniquely determined justified distinguished utility function. But in general, there may be more than one (see for instance [17]).

The weight values of the justified distinguished utility function are $k_1 = 1$ and $k_2 = 1$, as can easily be verified by solving the set of two constraints $+2k_1 + k_2 = 2 + 2k_1 - k_2$ and $+2k_1 + k_2 = 4 - 2k_1 + k_2$. The weight $k_i = 1$ is of course the typical value for non-redundant desires without constraint interaction.

$$\begin{array}{lll} u(w) = +2k_1 + k_2 & \text{if } w \models h \wedge w & (3) \\ \quad +2k_1 - k_2 & \text{if } w \models h \wedge \neg w & (1) \\ \quad -2k_1 + k_2 & \text{if } w \models \neg h \wedge w & (-1) \\ \quad -2k_1 - k_2 & \text{if } w \models \neg h \wedge \neg w & (-3) \end{array}$$

$$\begin{array}{l} \text{s.t.}: +2k_1 + k_2 \geq 2 + 2k_1 - k_2 \\ \quad +2k_1 + k_2 \geq 4 - 2k_1 + k_2 \end{array}$$

A consequence of the fact that the strength of 'to be healthy' is higher than the strength of 'to be wealthy' is the intuitive result that in the justified distinguished utility functions 'healthy and not wealthy' ($h \wedge \neg w$) is better than 'unhealthy and wealthy' ($\neg h \wedge w$).

To further illustrate the D_{MM} desires we now discuss the following extension of the specificity set in Example 3. It also illustrates how more specific desires override more general ones.

Example 26 (Specificity, continued). Consider the desire specification that consists of the following three loss desires with explicit strength 2. The second desire is more specific than the first one (as in Example 3), and, moreover, the third one is more specific than the second one.

$$\begin{array}{ll} D^1_{MM:2}(\neg s \mid \top) & \neg s \succ^1_{MM:2} s \\ & \quad Mu(\neg s) \geq 2 + Mu(s) \\ D^1_{MM:2}(s \mid i) & s \wedge i \succ^1_{MM:2} \neg s \wedge i \\ & \quad Mu(s \wedge i) \geq 2 + Mu(\neg s \wedge i) \\ D^1_{MM:2}(\neg s \mid i \wedge j) & \neg s \wedge i \wedge j \succ^1_{MM:2} s \wedge i \wedge j \\ & \quad Mu(\neg s \wedge i \wedge j) \geq 2 + \\ & \quad Mu(s \wedge i \wedge j) \end{array}$$

Again we consider the distinguished and justified distinguished utility functions. The weight values of the justified distinguished utility function are $k_1 = 1$, $k_2 = 2$, and $k_3 = 2$, the unique solution of the set of constraints $0 = 2 - 2k_1$, $-2k_1 = 2 - 2k_2$ and $-2k_2 = 2 - 2k_1 - 2k_3$.

$$\begin{array}{lll} u(w) = 0 & \text{if } w \models \neg s \wedge \neg i & (0) \\ \quad -2k_1 & \text{if } w \models s \wedge (\neg i \vee \neg j) & (-2) \\ \quad -2k_2 & \text{if } w \models \neg s \wedge i & (-4) \\ \quad -2k_1 - 2k_3 & \text{if } w \models s \wedge i \wedge j & (-6) \end{array}$$

$$\begin{array}{l} \text{s.t.}: \ 0 \geq 2 - 2k_1 \\ \quad -2k_1 \geq 2 - 2k_2 \\ \quad -2k_2 \geq 2 - 2k_1 - 2k_3 \end{array}$$

Now consider the change of polarity to a mixed gain-loss value 0.5, i.e. consider the desire specification

$$\left\{ D^{0.5}_{MM:2}(\neg s \mid \top),\ D^{0.5}_{MM:2}(s \mid i),\ D^{0.5}_{MM:2}(\neg s \mid i \wedge j) \right\}$$

The constraints remain the same as above, but the set of constructible utility functions changes. We give the distinguished and justified distinguished utility functions, the weight values of the latter being $k_1 = 5$, $k_2 = 8$, $k_3 = 4$.

$$\begin{array}{lll} u(w) = +k_1 & \text{if } w \models \neg s \wedge \neg i & (5) \\ \quad -k_1 + k_2 & \text{if } w \models s \wedge i \wedge \neg j & (3) \\ \quad +k_1 - k_2 + k_3 & \text{if } w \models \neg s \wedge i \wedge j & (1) \\ \quad -k_1 + k_2 - k_3 & \text{if } w \models s \wedge i \wedge j & (-1) \\ \quad +k_1 - k_2 & \text{if } w \models \neg s \wedge i \wedge \neg j & (-3) \\ \quad -k_1 & \text{if } w \models s \wedge \neg i & (-5) \end{array}$$

$$\begin{array}{l} \text{s.t.}: +k_1 \geq 2 - k_1 + k_2 \\ \quad -k_1 + k_2 \geq 2 + k_1 - k_2 + k_3 \\ \quad +k_1 - k_2 + k_3 \geq 2 - k_1 + k_2 - k_3 \end{array}$$

More specific desires override more general ones, because for justified distinguished u we have $u \not\models D(\neg s \mid i)$ and $u \not\models D(s \mid i \wedge j)$. The most notable distinction between the loss desires and the mixed gain-loss desires is the large gap in the latter between the worlds which are only affected by the first desire. With loss desires we have for $\neg s \wedge \neg i$ worlds w_1 and $s \wedge \neg i$ worlds w_2 that $u(w_1) - u(w_2) = 2$ or 6 and with mixed desires we have $u(w_1) - u(w_2) = 10$.

We now consider three transitivity examples. First, the extension of the transitivity set in Example 4 illustrates that loss desires and mixed loss-gain desires give the same distinguished utility functions, up to adding 3. Again the desirable behavior of Lang's original proposal is not lost by our generalization. The verification of these results is analogous to the examples above and left to the reader.

$$S_1 = \left\{ \begin{array}{l} D^1_{MM:2}(p \mid p \vee c), \\ D^1_{MM:2}(c \mid c \vee h), \\ D^1_{MM:2}(h \mid h \vee w) \end{array} \right\}$$

$$U_{JJ}(S_1) = \{u_c^{-2} + u_h^{-4} + u_w^{-6}\}$$

$$S_2 = \left\{ \begin{array}{l} D^{0.5}_{MM:2}(p \mid p \vee c), \\ D^{0.5}_{MM:2}(c \mid c \vee h), \\ D^{0.5}_{MM:2}(h \mid h \vee w) \end{array} \right\}$$

$$U_{JJ}(S_2) = \{u_p^{+3} + u_c^{+1} + u_h^{-1} + u_w^{-3}\}$$

Second, the following generalization of Example 11, based on the strongest desires D_{mM}, shows that the three conflict free sets *TRANS*, *CTD* and *ATD* are again equivalent, up to global shifting.

$$S_3 = \{D^{0.5}_{mM:2}(p \mid p \vee c), D^{0.5}_{mM:2}(c \mid c \vee h)\}$$

$$U_{JJ}(S_3) = \{u_p^{+2} + u_h^{-2}\}$$

$$S_4 = \{D^{0.5}_{mM:2}(p \mid p \vee c \vee h), D^{0.5}_{mM:2}(c \mid c \vee h)\}$$

$$U_{JJ}(S_4) = \{u_p^{+1.5} + u_c^{-0.5} + u_h^{-2.5}\}$$

$$S_5 = \{D^{0.5}_{mM:2}(p \mid p \vee c), D^{0.5}_{mM:2}(p \vee c \mid p \vee c \vee h)\}$$

$$U_{JJ}(S_5) = \{u_p^{+2.5} + u_c^{+0.5} + u_h^{-1.5}\}$$

Third, the so-called deliberating robber [25] in Example 27, also based on the strongest desires, illustrates how changing the strength or polarity parameter influences the constructions.

Example 27 (The deliberating robber). Consider the following two sets of two desires, with the background knowledge that the r_i are mutually exclusive and exhaustive. For readability we only give the representation with $\succ$ operators.

$$\begin{array}{lll} DS_1 & r_2 \wedge c \succ^1_{mM:1} r_1 \wedge c & r_4 \wedge \neg c \succ^1_{mM:2} r_3 \wedge \neg c \\ & & r_3 \wedge \neg c \succ^1_{mM:1} r_2 \wedge \neg c \\ & & r_2 \wedge \neg c \succ^1_{mM:1} r_1 \wedge \neg c \\ DS_2 & r_6 \wedge c \succ^1_{mM:1} r_5 \wedge c & r_4 \wedge \neg c \succ^1_{mM:2} r_3 \wedge \neg c \\ & r_5 \wedge c \succ^1_{mM:1} r_4 \wedge c & r_3 \wedge \neg c \succ^1_{mM:2} r_2 \wedge \neg c \\ & r_4 \wedge c \succ^1_{mM:1} r_3 \wedge c & r_2 \wedge \neg c \succ^1_{mM:2} r_1 \wedge \neg c \\ & r_3 \wedge c \succ^1_{mM:1} r_2 \wedge c & \\ & r_2 \wedge c \succ^1_{mM:1} r_1 \wedge c & \end{array}$$

DS_2 has a higher granularity for the cases in context c than DS_1. The justified utility functions are

$$U_{JJ}(DS_1) = \{u^{-1}_{r_1 \wedge c} + u^{-1}_{r_3 \wedge \neg c} + u^{-2}_{r_2 \wedge \neg c} + u^{-3}_{r_1 \wedge \neg c}\}$$

$$U_{JJ}(DS_2) = \{u^{-1}_{r_5 \wedge c} + u^{-2}_{r_4 \wedge c} + u^{-4}_{r_3 \wedge c} + u^{-5}_{r_2 \wedge c} + u^{-1}_{r_1 \wedge c} + u^{-1}_{r_3 \wedge \neg c} + u^{-2}_{r_2 \wedge \neg c} + u^{-3}_{r_1 \wedge \neg c}\}$$

The justified utility functions correspond to the System Z model of the desire specification. The problem of this example in System Z, as discussed extensively in [28], is that $D(c \mid r_1)$ is derived from DS_1 and $D(\neg c \mid r_1)$ from DS_2. This is counterintuitive, because:

- in DS_1 intuitively the least preferred worlds for context c are $c \wedge r_1$ worlds, and the most preferred worlds for context $\neg c$ are $\neg c \wedge r_1$ worlds, and these two sets of worlds are incomparable. Context c is preferred, because it contains only one violation of a desire, whereas the latter contains three violations.
- in DS_2 we have further specified the degrees of robbery in context c, we derive the desire to be in context $\neg c$ instead of context c. Hence, the system is sensitive to the granularity of the specification, which is counterintuitive.

With the parameterized desires we can change the construction by changing the strength or polarity parameter. For example, if we change the strength of the first desire of DS_1 to 3, then the $r_1 \wedge c$ worlds and the $r_1 \wedge \neg c$ worlds both have utility -3. Moreover, if we change all loss desires into gain desires, then the $r_1 \wedge c$ worlds and the $r_1 \wedge \neg c$ worlds both have utility 0.

The ad hoc adjustment of parameters in the previous example clearly shows the advantages and the drawbacks of our approach. The advantage is that we have a lot of control over the construction of the utility functions, but the drawback is that other problems may be expected when the parameters are changed due to its ad hoc character. For example, when we change to gain desires, then the best worlds are no longer equivalent—and this is in principle just as counterintuitive! We now have the tools to influence the construction but there are no general guidelines about the choice of parameters, this depends on the meaning of the propositions.

The following continuation of Example 13, based on the weakest desires, illustrates how changing the lifting parameter influences the constructions.

Example 28 (Marriage, continued). Consider the desire specification $DS_1 =$

$$\{D^1_{Mm:2}(j \mid \top), D^1_{Mm:2}(f \mid \top), D^1_{Mm:2}(\neg(j \wedge f) \mid \top)\}$$

The justified values are for $k_1 = 1, k_2 = 1, k_3 = 2$.

$$\begin{aligned} U_J &= \{u^{-2k_2}_{j\wedge\neg f} + u^{-2k_1}_{\neg j\wedge f} + u^{-2k_1-2k_2}_{\neg j\wedge\neg f} + u^{-2k_3}_{j\wedge f}\} \\ \text{s.t.:}\ & \max(-2k_2, -2k_3) \geq 2 - 2k_1 - 2k_2 \\ & \max(-2k_1, -2k_3) \geq 2 - 2k_1 - 2k_2 \\ & \max(-2k_1, -2k_2) \geq 2 - 2k_3 \\ U_{JJ} &= \{u^{-2}_{j\wedge\neg f} + u^{-2}_{\neg j\wedge f} + u^{-4}_{\neg j\wedge\neg f} + u^{-4}_{j\wedge f}\} \end{aligned}$$

Moreover, consider

$$DS_2 = \{D^{0.5}_{Mm:2}(j \mid \top), D^{0.5}_{Mm:2}(f \mid \top), D^{0.5}_{Mm:2}(\neg(j \wedge f) \mid \top)\}$$

Justification here obtains for $k_1 = 1, k_2 = 1, k_3 = 2$.

$$\begin{aligned} U_J &= \{u^{+k_1-k_2+k_3}_{j\wedge\neg f} + u^{-k_1+k_2+k_3}_{\neg j\wedge f} + u^{-k_1-k_2+k_3}_{\neg j\wedge\neg f} + u^{+k_1+k_2-k_3}_{j\wedge f}\} \\ \text{s.t.:}\ & \max(+k_1 - k_2 + k_3, +k_1 + k_2 - k_3) \geq 2 - k_1 - k_2 + k_3 \\ & \max(-k_1 + k_2 + k_3, +k_1 + k_2 - k_3) \geq 2 - k_1 - k_2 + k_3 \\ & \max(+k_1 - k_2 + k_3, -k_1 + k_2 + k_3) \geq 2 + k_1 + k_2 - k_3 \end{aligned}$$

$$U_{JJ} = \{u^2_{j\wedge\neg f} + u^2_{\neg j\wedge f} + u^0_{\neg j\wedge\neg f} + u^0_{j\wedge f}\}$$

If we think that being married to no-one is not equivalent to being married to both, then we can change the parameters of the last desire. For example, if we change the condition from *Mm* to *mM* then we get the following result. The distinguished utility functions are constructed as before, but the third constraint is replaced by the following new constraint.

$$\text{s.t.:} -k_1 - k_2 + k_3 \geq 2 + k_1 + k_2 - k_3$$

They induce the justified values $k_1 = 1$, $k_2 = 1$, $k_3 = 3$ and the corresponding justified utility function

$$U_{JJ} = \{u^3_{j\wedge\neg f} + u^3_{\neg j\wedge f} + u^1_{\neg j\wedge\neg f} + u^{-1}_{j\wedge f}\}$$

Thus, replacing the last desire by a stronger constraint induces the desired result that the situation in which Sue is married to both is the least desired state of all.

The following extension of Example 14, also based on the weakest desires, illustrates a construction in which some desires are redundant.

Example 29 (Fence and dog, continued). Consider the following desire specification $DS =$

$$\{D^{0.5}_{Mm:2}(\neg f \mid \top), D^{0.5}_{Mm:2}(f \mid d), D^{0.5}_{Mm:2}(d \mid \top)\}$$

The distinguished utility functions are constructed as follows, together with the constraints and distinguished justified values for $k_1 = 1, k_2 = 2, k_3 = 0$.

$$u(w) = \begin{cases} +k_1 - k_3 & \text{if } w \models \neg f \wedge \neg d \quad (1) \\ -k_1 + k_2 + k_3 & \text{if } w \models f \wedge d \quad (1) \\ +k_1 - k_2 + k_3 & \text{if } w \models \neg f \wedge d \quad (\text{-}1) \\ -k_1 - k_3 & \text{if } w \models f \wedge \neg d \quad (\text{-}1) \end{cases}$$

$$\begin{aligned} &\text{s.t.: } \max(+k_1 - k_3, +k_1 - k_2 + k_3) \geq 2 - k_1 - k_3 \\ &-k_1 + k_2 + k_3 \geq 2 + k_1 - k_2 + k_3 \\ &\max(-k_1 + k_2 + k_3, +k_1 - k_2 + k_3) \geq 2 - k_1 - k_3 \end{aligned}$$

The latter desire is redundant in the justified distinguished context, because $k_3 = 0$. If we think that our desire to have a dog is not well represented, then we can change the parameters of the desire. For example, we can replace the lifting policy of the desire by *MM*. The distinguished utility functions are constructed as before, but the third constraint is replaced by the following new constraint.

s.t.: $\max(-k_1+k_2+k_3, +k_1-k_2+k_3) \geq 2+k_1-k_3$

They induce the justified values $k_1=0$, $k_2=1$, $k_3=0.5$ and utilities for $\neg f \wedge d$, $f \wedge d$, $\neg f \wedge d$ and $f \wedge \neg d$ of respectively −0.5, 1.5, −0.5, −0.5. Thus, replacing the last desire by a stronger constraint induces the desired result that the situation in which there is a dog and a fence is the most desired state of all. The third desire is no longer redundant, but the first one is.

6. Conclusions

In this paper we have studied and extended the logic of desires in Lang's framework for qualitative decision theory. First, 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. We have shown how 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. Second, we have shown the formal relation between Lang's logic and Weydert's logic originally developed for defaults, and we have shown how extensions of the underlying so-called 'epistemic constructibility' can be used in the construction of distinguished utility functions. 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 other minimization principles, 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. Finally, a discussion of related work in [29, 30] will have to wait until we build up a full decision-theoretic framework in which utility and plausibility are combined.

Notes

1. 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. Both approaches 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. [26], 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.
2. In finite contexts, it is sufficient to know the values for singletons $R(\{w\})$, abbreviated by $R(w)$.
3. This name refers to Jeffrey-conditionalization, see also [31].
4. However, both systems satisfy a weaker notion of redundancy, they are non-sensitive to repetition. The two notions of redundancy are formally defined in Definition 17.
5. Another interesting possibility is to assume that all the worlds have the same probability and to compute the expected utility, which gives the average of the utilities. However, here we do not explore this alternative, because it constitutes a departure of our qualitative perspective.
6. Weydert has proven this for his defaults, i.e. for simple loss desires ($D^1_{MM:s}$). It is an open problem whether it can be proven in a more general context, e.g. for all $D^p_{MM:s}$ desires.

References

1. R.D. Luce and H. Raiffa, *Games and Decisions*, John Wiley: New York, 1957.
2. R. Jeffrey, *The Logic of Decision*, University of Chicago Press, 2nd ed. 1983.
3. R.L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-offs*, Wiley and Sons. New York, 1976.
4. J. Doyle and M.P. Wellman, "Preferential semantics for goals," in *Proceedings of AAAI-91*, Anaheim, 1991, pp. 698–703.
5. 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.
6. C. Boutilier, "Toward a logic for qualitative decision theory," in *Proceedings of the KR'94*, 1994, pp. 75–86.
7. J. Lang, "Conditional desires and utilities—an alternative approach to qualitative decision theory," in *Proceedings of the ECAI'96*, 1996, pp. 318–322.
8. J.D. Mullen, "Does the logic of preference rest on a mistake," *Metaphilosophy*, vol. 10, pp. 247–255, 1979.
9. J. Pearl, "From conditional ought to qualitative decision theory," in *Proceedings of the UAI'93*, 1993, pp. 12–20.
10. F. Bacchus and A.J. Grove, "Utility independence in a qualitative decision theory," in *Proceedings of KR'96*, 1996, pp. 542–552.
11. S.-W. Tan and J. Pearl, "Specification and evaluation of preferences under uncertainty," in *Proceedings of the KR'94*, 1994, pp. 530–539.
12. G. Pinkas, "Reasoning, nonmonotonicity and learning in connectionist networks that capture propositional knowledge," *Artificial Intelligence*, vol. 77, pp. 203–247, 1995.
13. Y. Shoham, *Reasoning About Change: Time and Causation from the Stand Point of Artificial Intelligence*, MIT Press, Cambridge, Massachusetts, 1988.

14. S. Kraus, D. Lehmann, and M. Magidor, “Nonmonotonic reasoning, preferential models and cumulative logics,” *Artificial Intelligence*, vol. 44, pp. 167–207, 1990.
15. E. Weydert, “Default entailment a preferential construction semanaics for defeasible inference,” in *Annual German Conference on Artificial Intelligence (KI-19): Bielefeld, Germany, September 11-13, 1995; proceedings*, edited by I. Wachsmuth, C.-R. Rollinger, and W. Brauer, Springer: Berlin, vol. LNAI 981, pp. 173–184, 1995.
16. E. Weydert, “System J—revision entailment: Default reasoning through ranking measure updates,” in *Practical Reasoning—International Conference on Formal and Applied Practical Reasoning, FAPR’96*, vol. 1085 of *Lecture Notes in Computer Science*, edited by D. Gabbay and H.J. Ohlbach, Springer: Bonn, Germany, pp. 637–649, June 1996.
17. E. Weydert, “System JZ: How to build a canonical ranking model of a default knowledge base,” in *Proceedings of the Seventh International Cconference on Knowledge Representation and Reasoning (KR’98)*, 1998, pp. 190–201.
18. E. Weydert, “General belief measures,” in *Tenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann: San Mateo, CA, 1994, pp. 575–582.
19. W. Spohn, “A general non-probabilistic theory of inductive reasoning,” in *Uncertainty in Artifical Intelligence 4 (UAI’90)*, North Holland: Amsterdam, 1990.
20. D. Dubois and H. Prade, *Possibility theory*. Plenum Press: New York, 1988.
21. R. Thomason and R. Horty, “Nondeterministic action and dominance: Foundations for planning and qualitative decision,” in *Proceedings of the TARK’96*, Morgan Kaufmann, 1996, pp. 229–250.
22. D. Lewis, *Counterfactuals*, Blackwell: Oxford, 1973.
23. L. van der Torre and Y. Tan, “Contrary-to-duty reasoning with preference-based dyadic obligations,” *Annals of Mathematics and Artificial Intelligence*, vol. 27, pp. 49–78, 2000.
24. H. Prakken and M.J. Sergot, “Contrary-to-duty obligations,” *Studia Logica*, vol. 57, pp. 91–115, 1996.
25. J. Doyle, Y. Shoham, and M.P. Wellman, “The logic of relative desires” in *Sixth International Symposium on Methodologies for Intelligent Systems*, Charlotte, North Carolina, 1991.
26. S.-W. Tan and J. Pearl, “Qualitative decision theory,” in *Proceedings of the AAAI’94*, 1994.
27. 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)*, 1992, pp. 236–242.
28. L. van der Torre and Y. Tan, “Deliberate robbery, or the calculating samaritan,” in *Proceedings of the ECAI’98 Workshop on Practical Reasoning and Rationality (PRR’98)*, 1998.
29. R.I. Brafman and N. Friedman, “On decision-theoretic foundations for defaults,” in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, edited by C.S. Mellish, Morgan Kaufmann, pp. 1458–1465, 1995.
30. B. Bonet and H. Geffner, “Arguing for decisions: A qualitative model of decision making,” in *Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence (UAI-96)*, edited by E. Horvitz and F. Jensen, Morgan Kaufmann Publishers, pp. 98–105, 1996.
31. M. Goldszmidt and J. Pearl, “Qualitative probabilities for default reasoning, belief revision, and causal modeling,” *Artificial Intelligence*, vol. 84, pp. 57–112, 1996.

Leendert van der Torre is a researcher at the Department of Artificial Intelligence of the Vrije Universiteit in Amsterdam. He obtained a PhD degree from the University of Rotterdam (the Netherlands) on “Reasoning about Obligations: Defeasibility in Preference-based deontic Logic” and afterwards he was a visiting researcher at the Max Planck Institute for Computer Science (Germany) and the University of Toulouse (France). His research interests are in formal approaches to practical reasoning, in particular normative reasoning (deontic logic) and decision making (qualitative decision theory). http://www.cs.vu.nl

Emil Weydert has studied mathematics, logic, and computer science at the University of Bonn, where he received his PhD in 1988. He worked as a researcher, first at the Institute for Computational Linguistics in Stuttgart, then. since 1993, at the Max-Planck-Institute for Computer Science in Saarbrücken. His current research interests include nonmonotonic and probabilistic reasoning, as well as autonomous agents. http://www.mpi-sb.mpg.de