Permissions and Uncontrollable Propositions in DSDL3: Non-Monotonicity and Algorithms

Souhila Kaci[1] and Leendert van der Torre[2]

[1] Centre de Recherche en Informatique de Lens (C.R.I.L.)–C.N.R.S, France
[2] University of Luxembourg

Abstract. In this paper we are interested in non-monotonic extensions of Bengt Hansson's standard dyadic deontic logic 3, known as DSDL3. We study specificity principles for DSDL3 with both controllable and uncontrollable propositions. We introduce an algorithm for minimal specificity which not only covers obligations but also permissions, and we discuss the distinction between weak and strong permissions. Moreover, we introduce ways to combine algorithms for minimal and maximal specificity for DSDL3 with controllable and uncontrollable propositions, based on 'optimistic' and 'pessimistic' reasoning respectively.

1 Introduction

Hansson's standard dyadic deontic logic 3 [9], known as DSDL3, is an extension of standard deontic logic, SDL, also known as system KD, with dyadic obligations. It has been called a defeasible deontic logic because it does not satisfy unrestricted strengthening of the antecedent, the derivation of $O(p|q \wedge r)$ from $O(p|q)$. Spohn's axiom in his axiomatization of DSDL3 [18] informs us that strengthening of the antecedent only holds conditional to a permission, where $P(p|q) = \neg O(\neg p|q)$:

$$P(r|q) \rightarrow (O(r \rightarrow p|q) \leftrightarrow O(p|q \wedge r))$$

Monotonic and non-monotonic extensions to DSDL3 have been studied to strengthen the antecedent. The former has been studied using notions of settledness or necessity by, for example, Prakken and Sergot [16]. The latter has been directly inspired by the interpretation of DSDL3 as a theory of default conditionals, or more generally as a framework for non-monotonic logic following the work of Shoham [17] and Kraus, Lehmann and Magidor [11]. The main approach in this setting to strengthen the antecedent is based on the so-called minimal specificity principle by, amongst others, Lehmann and Magidor [13] and Boutilier [4]. These non-monotonic extensions are accompanied by efficient algorithms [15], though these algorithms have the drawback to be defined only for sets of dyadic obligations, not for more complex formulae such as, for example, permissions.

DEON2006 has a special focus on deontic notions in the theory, specification and implementation of artificial normative systems, such as electronic institutions, norm-regulated multi-agent systems, and artificial agent societies more generally. In the context of agent theory, Boutilier studies non-monotonic DSDL3 extended with the distinction between controllable and uncontrollable propositions [5]. Though this distinction originates from the areas of discrete event systems and control theory, Boutilier uses it as a simple theory of decision (or action) in qualitative decision theory. It has been further developed by, for example, Lang *et al.* [12] and Cholvy and Garion [8].

In this paper we are interested in the following questions:

1. How can we extend non-monotonic DSDL3 with permissions?
2. What is the relevance of the distinction between controllable and uncontrollable propositions in non-monotonic DSDL3?

Despite the work in non-monotonic extensions of DSDL3 for default conditionals [4, 1], desires [12], and preferences [2, 10] in artificial intelligence, somewhat surprisingly the extension of existing algorithms to permission seems to have received less attention. Apparently, whereas permission plays a central role in deontic logic, the analogous negation of default conditionals, absence of desires, and non-strict preference are of less interest in the other research areas. Though there are related extensions, such as ones dealing with equalities, the only extension of algorithms concerned with permission we are aware of has been proposed by Booth and Paris [3]. However, their algorithm is inefficient as it requires the construction of a potentially large number of pre-orders. Our algorithm constructs only the minimal specific pre-order.

When a distinction between controllable and uncontrollable propositions is introduced in DSDL3, one may revisit the use of the minimal specificity principle. For example, another option would be to use the *maximal* specificity principle, which does not assume that worlds are as normal as possible, or gravitate towards the ideal, but which assumes that worlds are as abnormal as possible, or gravitate towards the worst. We argue that whereas the 'optimistic' reasoning underlying the minimal specificity principle may make sense for controllable propositions, because, for example, any rational agent will see to it that the best state will be realized, for uncontrollable propositions a more 'pessimistic' attitude may be used as well. We also study the combination of both kinds of reasoning.

In this paper we do not discuss the advantages and disadvantages of DSDL3, nor of non-monotonic DSDL3, since they have already been discussed extensively during the last 35 years. For the same reason we do not present the usual examples again, but we focus on the logical properties of the system. Next to SDL, DSDL3 is probably the best known deontic logic, and the most successful logic developed in deontic logic and used outside this area (in particular in artificial intelligence). We also do not discuss its well known relation to preference logic, due to the fact that "the best q are p" is equivalent to "$p \wedge q$ is preferred to $\neg p \wedge q$" in several preference logics, see for example [20]. However, we believe that the preference-based reading of DSDL3 suggests that the 'optimistic' reading of non-monotonic DSDL3 may be arbitrary, because in this representation, there does not seem to be a reason why we compare the best $p \wedge q$ worlds, and not the worst ones (or, for example, the ones in the middle).

The layout of this paper is as follows. In Section 2 we repeat the definitions of non-monotonic DSDL3, and in Section 3 we present the algorithm to compute the most specific pre-order satisfying a set of obligations and permissions. In Section 4 we repeat the distinction between controllable and uncontrollable propositions, and we present the algorithms for the uncontrollable case. In Section 5 we consider the merging of the two kinds of obligations.

2 Non-monotonic extension of DSDL3

Norm specifications consist of obligations and permissions. $O(p|q)$ is read as 'p is obligatory if q' and $P(p|q)$ is read as 'p is permitted if q'.

Definition 1 (Norm specification). *Let $\mathcal{A}$ be a finite set of propositional atoms, and $\mathcal{L}$ a propositional logic based on $\mathcal{A}$. A norm specification is a set of norms $\mathcal{C} = \mathcal{C}_O \cup \mathcal{C}_P$ where for $p_i, q_i, p'_j, q'_j \in \mathcal{L}$:*

$$\mathcal{C}_O = \{C_i = O(p_i|q_i) \mid i = 1 \dots n\}$$

$$\mathcal{C}_P = \{C'_j = P(p'_j|q'_j) \mid j = 1 \dots m\}$$

The norms are interpreted on a total pre-order on the propositional valuations (or worlds).

Definition 2 (Monotonic semantics). *Let $\mathcal{A}$ and $\mathcal{L}$ be as before, let "worlds" W be the set of propositional valuations of $\mathcal{L}$, and $\succeq$ a total pre-order on W. Let $|\phi|$ be the set of propositional models of ϕ. We write $w \succ w'$ for $w \succeq w'$ without $w' \succeq w$, and we write $\max(p, \succeq)$ for $\{w \in |p| \mid \forall w' \in |p|$ we have $w \succeq w'\}$. Satisfiability is defined as follows:*

$\langle W, \succeq\rangle \models O(p|q)$ *iff* $\max(q, \succeq) \subseteq |p|$, *which is equivalent to stating that* $\forall \omega \in \max(p \wedge q, \succeq), \forall \omega' \in \max(p \wedge \neg q, \succeq)$, *we have* $\omega \succ \omega'$.

Moreover, we define

$\langle W, \succeq\rangle \models P(p|q)$ *iff* $\langle W, \succeq\rangle \models \neg O(\neg p|q)$ *which is equivalent to stating that* $|p \wedge q| \neq \emptyset$, *and* $\forall \omega \in \max(p \wedge q, \succeq), \forall \omega' \in \max(p \wedge \neg q, \succeq)$, *we have* $\omega \succeq \omega'$.

A total pre-order $\succeq$ is a model of (satisfies) a norm specification $\mathcal{C}$ iff it satisfies each norm in the specification $\mathcal{C}$. We write $\mathcal{M}(\mathcal{C})$ for the set of models of $\mathcal{C}$.

For an infinite set of propositional atoms $\mathcal{A}$, a more sophisticated definition proposed by Lewis [14] and popularized in AI by Boutilier [4], deals with infinite descending chains. They define $\langle W, \succeq\rangle \models O(p|q)$ iff $|q| = \emptyset$, or there exists a $p \wedge q$ world w such that there does not exist a $\neg p \wedge q$ world w' with $w' \geq w$.

In the algorithm we do not use the total pre-order $\succeq$ directly, but we use an equivalent representation as an ordered partition, defined as follows. E_1 is the set of ideal worlds, and E_n is the set of worst worlds.

Definition 3 (Ordered partition). *A sequence of sets of worlds of the form $(E_1, \cdots, E_n)$ is an ordered partition of W iff $\forall i$, E_i is nonempty, $E_1 \cup \cdots \cup E_n = W$ and $\forall i, j$, $E_i \cap E_j = \emptyset$ for $i \neq j$. An ordered partition of W is associated with pre-order $\succeq$ on W iff $\forall \omega, \omega' \in W$ with $\omega \in E_i, \omega' \in E_j$ we have $i \leq j$ iff $\omega \succeq \omega'$.*

In this section we compare total pre-orders based on the so-called *minimal specificity principle* which is also known as System Z or gravitating towards the ideal.

Definition 4 (Preference semantics). *Let $\succeq$ and $\succeq'$ be two total pre-orders on a set of worlds W represented by ordered partitions $(E_1, \cdots, E_n)$ and $(E'_1, \cdots, E'_m)$ respectively. We say that $\succeq$ is at least as specific as $\succeq'$, written as $\succeq' \sqsubseteq \succeq$, iff $\forall \omega \in W$, if $\omega \in E_i$ and $\omega \in E'_j$ then $i \leq j$. $\succeq$ is less specific as $\succeq'$, written as $\succeq \sqsubset \succeq'$, iff $\succeq \sqsubseteq \succeq'$ without $\succeq' \sqsubseteq \succeq$. $\succeq$ is the least specific pre-order among a set of pre-orders $\mathcal{O}$ if there is no $\succeq'$ in $\mathcal{O}$ such that $\succeq' \sqsubset \succeq$.*

The following example illustrates minimal specificity.

Example 1. Consider the single obligation $O(p|q)$. Applying the minimal specificity principle gives the following model $\succeq = (|p \wedge q| \cup |p \wedge \neg q| \cup |\neg p \wedge \neg q|, |\neg p \wedge q|)$. The ideal worlds in this model are those which do not violate the obligation. More precisely, worlds in $|p \wedge q|$ belong to the set of ideal worlds since they fulfill the obligation, but worlds in $|p \wedge \neg q|$ and $|\neg p \wedge \neg q|$ are ideal too since they do not violate the rule even if they do not fulfill it.

Shoham [17] defines non-monotonic consequences of a logical theory as all formulas which are true in the 'preferred' models of the theory. An attractive property is case is which there is only one 'preferred' model, because in that case it can be decided whether a formula non-monotonically follows from a logical theory by calculating the unique 'preferred' model, and testing whether the formula is satisfied by the 'preferred' model. Likewise, finding all non-monotonic consequences can be found by calculating the unique 'preferred' model and characterizing all formulas satisfied by this model.

Definition 5 (Non-monotonic entailment). *A norm specification C preferentially implies $O(p|q)$ (or $P(p|q)$) if and only if for least specific models of C are also a model of $O(p|q)$ (or $P(p|q)$).*

The following example illustrates non-monotonic entailment, which can be used to reason about violations or exceptions.

Example 2 (Continued). The norm specification consisting of the obligation $O(p|q)$ preferentially implies $O(p|q \wedge r)$, but the norm specification consisting of both $O(p|q)$ and $O(\neg p|q \wedge r)$ does not preferentially imply $O(p|q \wedge r)$.

3 Algorithm for obligations and permissions

The algorithm to calculate the least specific pre-order of a norm specification is given in Algorithm 1.1. The basic idea of the algorithm is to construct the least specific pre-order by calculating the sets of worlds of the ordered partition, going from ideal to the worst worlds. It extends the known algorithm [15, 2] for obligations with the second line to check whether the individual permissions are consistent, the inner while loop, to deal with permissions, and the second removal clause in the end to take care of the removal of permissions. Given a norm specification, let $\mathcal{C} = \mathcal{C}_O \cup \mathcal{C}_P$ where

$$\mathcal{C}_O = \{C_i : O(p_i|q_i)|i = 1 \cdots n\} \text{ and } \mathcal{C}_P = \{C'_j : P(p'_j|q'_j)|j = 1 \cdots m\}$$

and let $\mathcal{L} = \{(L(C_i), R(C_i)) : C_i \in \mathcal{C}_O\} \cup \{(L(C'_j), R(C'_j)) : C'_j \in \mathcal{C}_P\}$, where $L(C_i) = |p_i \wedge q_i|$, $R(C_i) = |\neg p_i \wedge q_i|$, $L(C'_j) = |p'_j \wedge q'_j|$ and $R(C'_j) = |\neg p'_j \wedge q'_j|$.

Algorithm 1.1: Handling obligations and permissions.

begin
  if *any* $L(C'_i) = \emptyset$ **then** Stop (inconsistent constraints);
  $m = 0$; $W =$ set of all models of $\mathcal{L}$;
  while $W \neq \emptyset$ **do**
    – $m \leftarrow m + 1$, $i = 1$;
    – $E_m = \{\omega : \forall (L(C_i), R(C_i)) \in \mathcal{L}_\mathcal{C}, \omega \notin R(C_i)\}$;
    while $i = 1$ **do**
      i=0;
      for *each* $(L(C'_j), R(C'_j)$ *in* $\mathcal{L}_\mathcal{C}$ **do**
        if *(*$L(C'_j) \cap E_m = \emptyset$ *and* $R(C'_j) \cap E_m \neq \emptyset$*)* **then** $E_m = E_m \backslash R(C'_j)$; i=1;
    – **if** $E_m = \emptyset$ **then** Stop (inconsistent constraints);
    – $W = W - E_m$;
    – remove from $\mathcal{L}_\mathcal{C}$ each $(L(C_i), R(C_i))$ such that $L(C_i) \cap E_m \neq \emptyset$;
    – remove from $\mathcal{L}_\mathcal{C}$ each $(L(C'_j), R(C'_j))$ such that $L(C'_j) \cap E_m \neq \emptyset$;
  return $(E_1, \cdots, E_m)$
end

If we consider the case without permissions, then the algorithm calculates the next equivalence class of the partitioning E_m by taking all worlds which do not violate one of the obligations. Once an obligation is satisfied by an equivalence class, it no longer constrains the construction of the preorder, and can be removed.

With permissions, the construction is complicated since we cannot directly define the equivalence class E_m. The definition of E_m in line 6 of the algorithm is therefore an upper bound of this class. To make sure that all permissions are satisfied, thereafter some worlds may have to be removed from E_m. Moreover, once some worlds are removed, it may be the case that permissions which were already checked are now violated, so we have to reconsider them too (for which we use the variable j). Removal of permissions is analogous to the removal of obligations.

In the remainder of this section, we prove that the algorithm calculates the least specific pre-order.

Lemma 1. *The total pre-order computed by algorithm 1 belongs to the set of least specific pre-orders of* $\mathcal{C}$*.*

Proof. This can be checked by construction. Since the set of constraints is finite, the algorithm terminates. Since E_m *cannot be the empty set, the sequence is an ordered partition. Let* $\succeq = (E_1, \cdots, E_n)$ *be this total pre-order. Suppose that* $\succeq$ *doesn't belong to the set of least specific pre-orders of* $\mathcal{C}$*, i.e., for some* $\omega \in E_j$ *we could have put* ω *in* E_i *with* $i < j$*. However* $\omega \notin E_j$ *because either:*

obligations in $\mathcal{C}_O$**.** $\omega \in E_j$ *means that* ω *falsifies obligations in* $\mathcal{C}_O$ *which are not falsified by worlds in* E_i *with* $i < j$*. So if we put* ω *in* E_i *with* $i < j$*, we get a contradiction,*

permissions in $\mathcal{C}_P$**.** *Following the algorithm,* $\omega \in E_i$ *because otherwise there is some permission* $P(p'|q')$ *in* $\mathcal{C}_P$ *for which the best worlds of* $p' \wedge q'$ *are in* E_k *and the best worlds of* $\neg p' \wedge q'$ *are in* E_l *with* $l < k$ *which is a contradiction.*

To show the uniqueness of the least specific pre-order of $\mathcal{C}$, we follow the line of the proof given in [1]. We first define the maximum of two preference orders.

Definition 6. *Let $\succeq$ and $\succeq'$ be two preference orders represented by their well ordered partitions $(E_1, \cdots, E_n)$ and $(E'_1, \cdots, E'_m)$ respectively. We define the $\mathcal{MAX}$ operator by $\mathcal{MAX}(\succeq, \succeq') = (E''_1, \cdots, E''_{min(n,m)})$, such that $E''_1 = E_1 \cup E'_1$ and $E''_k = (E_k \cup E'_k) - (\bigcup_{i=1,\cdots,k-1} E''_i)$ for $k = 2, \cdots, min(n, m)$, and the empty sets E''_k are eliminated by renumbering the non-empty ones in sequence.*

Lemma 2 proves the uniqueness of the least specific pre-order in $\mathcal{M}(\mathcal{C})$.

Lemma 2. *If there is a minimal specific pre-order, then it is unique.*

Proof. We first show that $\mathcal{MAX}(\succeq, \succeq') \in \mathcal{M}(\mathcal{C})$ (1). Let $\succeq = (E_1, \cdots, E_h)$, $\succeq' = (E'_1, \cdots, E'_{h'})$, $\succeq'' = (E''_1, \cdots, E''_{min(h,h')})$, and $P(p|q) \in \mathcal{C}$. $\succeq, \succeq' \in \mathcal{M}(\mathcal{C})$, i.e., $\succeq \models P(p|q)$ and $\succeq' \models P(p|q)$. In other words, $\max(p \wedge q, \succeq) \subseteq E_i$ and $\max(p \wedge \neg q, \succeq) \subseteq E_j$ such that $i \leq j$ and $\max(p \wedge q, \succeq') \subseteq E'_k$ and $\max(p \wedge \neg q, \succeq') \subseteq E'_l$ such that $k \leq l$. Following Definition 6, $\max(p \wedge q, \succeq'') \subseteq E''_{\min(i,k)}$ and $\max(p \wedge \neg q, \succeq'') \subseteq E''_{\min(j,l)}$. Since $i \leq j$ and $k \leq l$ we have $\min(i, k) \leq \min(j, l)$. We conclude $\succeq'' \models P(p|q)$. The proof for $O(p|q)$ is analogous and can be found in [1]. Consequently, $\mathcal{MAX}(\succeq, \succeq') \in \mathcal{M}(\mathcal{C})$.

Moreover, we have that $\mathcal{MAX}(\succeq, \succeq')$ is less specific than or identical to both $\succeq$ and $\succeq'$ (2), the proof can be found also in [1].

Finally, we prove that the lemma follows from the two items by contradiction. So suppose that there are two distinct minimal specific orders $\succeq$ and $\succeq'$. Then according to item (1), $\mathcal{MAX}(\succeq, \succeq')$ is also a model of the preference specification and according to item (2), it is less specific than either $\succeq$ or $\succeq'$. Contradiction.

We can now conclude:

Theorem 1. *Algorithm 1 computes the least specific model of $\mathcal{M}(\mathcal{C})$.*

Proof. Following Lemma 1 it computes a preference order which belongs to the set of the least specific models and following Lemma 2, this preference order is unique.

4 Ought-to-be and ought-to-do

Some approaches introduce a full fledged logic of actions in theories of rational decision, but Boutilier [5] introduces the distinction between controllable and uncontrollable propositions from discrete event systems and control theory in his qualitative decision theory. This relatively simple approach to actions has reached some popularity, see [12, 7, 19]. The reason is that this abstract representation of actions – which are typically called decision variables – lets us focus on other aspects of decision making than the usual issues concerning causality, frame axioms, etc.

In the context of deontic logic, the distinction between controllable and uncontrollable propositions can be used as a simple way to distinguish and study the relation between ought-to-be and ought-to-do obligations. Consider a dynamic deontic logic. Dynamic logic contains expressions like $[\alpha]p$, which can be read as 'after doing or executing α, p holds, and dynamic deontic logic contains expressions $O(\alpha)$, expressing an ought-to-do obligation for α, and $O(p)$, expressing an ought-to-be obligation for p. Now assume that we add propositions $done(\alpha)$ for every action statement α, together with the axiom $[\alpha]done(\alpha)$. In that case, we may say that $O(done(\alpha))$ is a kind of ought-to-do obligation. Summarizing, if we have $O(p)$ – which is short for $O(p|\top)$ for any tautology $\top$ – for uncontrollable p, then we may call it an ought-to-be obligation, and if we have $O(x)$ for controllable x, then we may call it an ought-to-do obligation.

Having made the distinction between the two kinds of obligations, we are now faced by the question whether their logic is distinct. Neither Boutilier nor the other researchers working on controllable and uncontrollable propositions seem to have introduced distinct logics or distinct non-monotonic extensions to represent the two kinds of obligations.

When we consider DSDL3 and the related minimal specificity principle, we may observe that both of them are 'optimistic', in the following sense. First, the logic of $O(p|q)$ only considers the best or ideal worlds. Second, the non-monotonic extension of the minimal specificity principle assumes that each world is as good as possible. But why not select a more 'pessimistic' approach? Note that the notion of 'optimistic' and 'pessimistic' should be read metaphorically, referring to psychological or decision-theoretic notions.

For controllable propositions, this choice seems justified to us. The agent can control the truth value of the propositions, and therefore he or she should see to it that the best world will be realized. But for uncontrollable propositions, it is less clear. the choice of the best worlds seems rather arbitrary. Moreover, in decision making, there is often a trend to reason pessimistically about the environment.

Therefore, in the remainder of this paper we study 'pessimistic' kinds of reasoning for ought-to-be obligations. The 'pessimistic' alternatives are that $O(p|q)$ no longer means that the best q worlds are p worlds, but that the worst q worlds are $\neg p$ worlds. Moreover, instead of assuming that worlds are as good as possible, we assume that worlds are as bad as possible. As one may expect, the 'pessimistic' definition and the 'pessimistic' specificity principle go well together.

From now on, we write O^+ and P^+ to refer to the usual kinds of 'optimistic' obligations and permissions of DSDL3, as studied thus far in this paper. Moreover, we introduce new 'pessimistic' obligations and permissions, which we write as O^- and P^-.

Definition 7 (Norm specification). *Let $\mathcal{C}$ and $\mathcal{U}$ be two disjoint finite sets of controllable resp. uncontrollable propositional atoms, and $\mathcal{L}$ a propositional logic based on $\mathcal{C} \cup \mathcal{U}$. A norm specification is a set of norms $\mathcal{C} = \mathcal{C}_O^+ \cup \mathcal{C}_P^+ \cup \mathcal{C}_O^- \cup \mathcal{C}_P^-$, where the 'optimistic' $mathcal C_O^+ \cup \mathcal{C}_P^+$ are defined using C only, and $\mathcal{C}_O^- \cup \mathcal{C}_P^-$ using U only.*

As before, the norms are interpreted on a total pre-order. The semantics are straightforward.

Definition 8 (Monotonic semantics). *Satisfiability is defined as follows:*

$\langle W, \succeq \rangle \models O^+(p|q)$ *iff* $\max(q, \succeq) \subseteq |p|$
$\langle W, \succeq \rangle \models O^-(p|q)$ *iff* $\min(q, \succeq) \subseteq |\neg p|$

Moreover, we define

$\langle W, \succeq \rangle \models P^+(p|q)$ *iff* $\langle W, \succeq \rangle \models \neg O^+(\neg p|q)$
$\langle W, \succeq \rangle \models P^-(p|q)$ *iff* $\langle W, \succeq \rangle \models \neg O^-(\neg p|q)$

The non-monotonic semantics based on maximal specificity principle are straightforward too, and the maximal specificity algorithm is simply the dual of the minimal specificity algorithm. The basic idea of the algorithm is to construct the most specific pre-order by calculating the sets of worlds of the ordered partition, going from worst to the ideal worlds. As can easily be verified, we obtained algorithm 2 by replacing left hand side and right hand side in various places. Moreover, the pre-order is constructed from worst to ideal class, so in the last line we have to reverse the order of the classes. Let $\mathcal{C} = \mathcal{C}_O \cup \mathcal{C}_P$ where

$$\mathcal{C}_O = \{C_i : O(p_i|q_i)|i = 1 \cdots n\} \text{ and } \mathcal{C}_P = \{C'_j : P(p'_j|q'_j)|j = 1 \cdots m\}$$

We put $\mathcal{L} = \{(L(C_i), R(C_i)) : C_i \in \mathcal{C}_O\} \cup \{(L(C'_j), R(C'_j)) : C'_j \in \mathcal{C}_P\}$, where $L(C_i) = |p_i \wedge q_i|$, $R(C_i) = |\neg p_i \wedge q_i|$, $L(C'_j) = |p'_j \wedge q'_j|$ and $R(C'_j) = |\neg p'_j \wedge q'_j|$.

Algorithm 1.2: Handling ought-to-be obligations and permissions.

begin
  if *any* $R(C'_i) = \emptyset$ **then** Stop (inconsistent constraints);
  $m = 0$; $W =$ set of all models of $\mathcal{L}$;
  while $W \neq \emptyset$ **do**
    – $m \leftarrow m + 1, i = 1$;
    – $E_m = \{\omega : \forall (L(C_i), R(C_i)) \in \mathcal{L}_\mathcal{C}, \omega \notin L(C_i)\}$;
    while $i = 1$ **do**
      i=0;
      for *each* $(L(C'_j), R(C'_j)$ *in* $\mathcal{L}_\mathcal{C}$ **do**
        if *(*$L(C'_j) \cap E_m \neq \emptyset$ *and* $R(C'_j) \cap E_m = \emptyset$*)* **then** $E_m = E_m \backslash L(C'_j)$; i=1;
    – **if** $E_m = \emptyset$ **then** Stop (inconsistent constraints);
    – $W = W - E_m$;
    – remove from $\mathcal{L}_\mathcal{C}$ each $(L(C_i), R(C_i))$ such that $R(C_i) \cap E_m \neq \emptyset$;
    – remove from $\mathcal{L}_\mathcal{C}$ each $(L(C'_j), R(C'_j))$ such that $R(C'_j) \cap E_m \neq \emptyset$);
  return $(E'_1, \cdots, E'_l)$ *s.t.* $\forall 1 \leq h \leq l, E'_h = E_{l-h+1}$
end

The 'pessimistic' approach may be criticized, just as we have criticized the 'optimistic' approach, and more sophisticated approaches may be developed. However, our more general point is that one may reason in a different way with ought-to-be and ought-to-do obligations, or with controllable and uncontrollable propositions - where the above suggestion is just an instance of that general idea. Any approach that deals with these two kinds of obligations in a different way has to solve the problem we address in the following section: how can these approaches be combined?

5 Merging ought-to-be and ought-to-do

Figure 1 visualizes our approach to combine two distinct ways to reason with ought-to-be and ought-to-do in DSDL3. Our approach is based on 'optimistic' reasoning about controllables, and 'pessimistic' reasoning about uncontrollables. The former represent the agent's rationality to choose the optimal state if he has the power to do so, and the latter represents Wald criterion: the decision-maker selects that strategy which is associated with the best possible worst outcome.

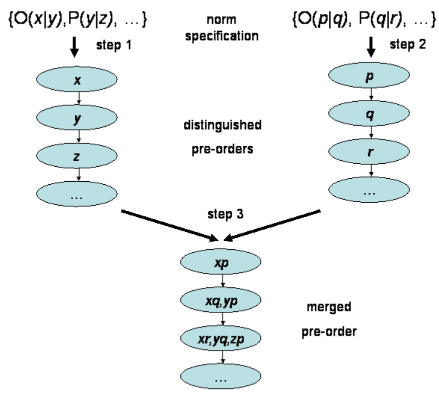

Fig. 1. Combining 'optimistic' and 'pessimistic' norms.

The norms in the norm specifications are interpreted as constraints on total pre-orders on worlds. Moreover, there are non-monotonic reasoning mechanisms to calculate distinguished pre-orders from the norm specifications. There is an 'optimistic' algorithm to calculate the unique distinguished total pre-order from the 'optimistic' norm specification (step 1), and a 'pessimistic' algorithm to calculate the unique distinguished total pre-order from the 'pessimistic' norm specification (step 2). Since we need a single preference order for decision making, we need to merge the two total pre-orders (step 3). We distinguish symmetric and a-symmetric mergers.

In this section we consider the merger of the least specific pre-order satisfying the 'optimistic' norm specification, and the most specific pre-order satisfying the 'pessimistic' norm specification. From now on, let $\mathcal{L}$ be a propositional language on disjoint sets of controllable and uncontrollable propositional atoms $\mathcal{C} \cup \mathcal{U}$. A norm specification consists of an 'optimistic' and a 'pessimistic' norm specification, i.e., 'optimistic' norms on controllables, and 'pessimistic' norms on uncontrollables. In general, let $\succeq$ be the merger of a pre-order $\succeq_d$ generated by 'optimistic' reasoning about ought-to-do obligations and a pre-order $\succeq_b$ generated by 'pessimistic' reasoning about ought-to-be obligations. We assume that the following conditions hold, known in economic theory as Arrows' conditions:

Definition 9. *Let $\succeq_d$, $\succeq_b$ and $\succeq$ be three total-preorders on the same set. $\succeq$ is a merger of $\succeq_d$ and $\succeq_b$ if and only if the following three conditions hold:*
If $w_1 \succ_d w_2$ and $w_1 \succ_b w_2$ then $w_1 \succ w_2$
If $w_1 \succ_d w_2$ and $w_1 \succeq_b w_2$ then $w_1 \succeq w_2$
If $w_1 \succeq_d w_2$ and $w_1 \succ_b w_2$ then $w_1 \succeq w_2$

Given two arbitrary pre-orders, there are many possible mergers. We therefore again consider distinguished pre-orders below.

The two minimal and maximal specific pre-orders of 'optimistic' and 'pessimistic' preference specifications satisfy the property that no two sets are disjoint.

Proposition 1. *Let $(E_1, \cdots, E_n)$ and $(E'_1, \cdots, E'_m)$ be the ordered partitions of $\succeq_d$ and $\succeq_b$ respectively. We have for all $1 \leq i \leq n$ and all $1 \leq j \leq m$ that $E_i \cap E'_j \neq \emptyset$.*
Proof. Due to the fact that $\succeq_d$ and $\succeq_b$ are defined on disjoint sets of variables.

The least and most specific mergers (thus satisfying Arrow's conditions) are unique and identical, and can be obtained as follows. Given Proposition 1, thus far nonempty sets E''_k do not exist, but we prefer a more general definition which can also be used in other mergers.

Proposition 2. *Let $(E_1, \cdots, E_n)$ and $(E'_1, \cdots, E'_m)$ be the ordered partitions of $\succeq_d$ and $\succeq_b$ respectively. The most specific merger of $\succeq_d$ and $\succeq_b$ is $\succeq = (E''_1, \cdots, E''_{n+m})$ such that if $\omega \in E_i$ and $\omega \in E'_j$ then $\omega \in E''_{i+j-1}$, and by eliminating nonempty sets E''_k and renumbering the non-empty ones in sequence.*

The most specific merger is illustrated by the following example.

Example 3. Consider the 'optimistic' preference specification $O(p)$ and the 'pessimistic' preference specification $O(m)$, where p and m stand respectively for "I have to work on a project in order to get money" and "my boss has given me money to pay my conference fee".
We have $\succeq_d = (\{mp, \neg mp\}, \{m\neg p, \neg m\neg p\})$ and $\succeq_b = (\{mp, m\neg p\}, \{\neg mp, \neg m\neg p\})$.
The most specific merger is $\{\{mp\}, \{\neg mp, m\neg p\}, \{\neg m\neg p\}\}$.

Analogously we may also consider the product rule (...then $\omega \in E''_{i*j}$), or other symmetric mergers.

The minimax merger gives priority to the preorder associated to the 'optimistic' preference specification and computed following the minimal specificity principle ($\succeq_d$) over the one associated to the 'pessimistic' preference specification and computed following the maximal specificity principle ($\succeq_b$). Indeed alternatives are first ordered w.r.t. $\succeq_d$ and only in the case of equality $\succeq_b$ is considered.

Definition 10 (Minimax merger). *$w_1 \succ w_2$ iff $w_1 \succ_1 w_2$ or $w_1 \sim_1 w_2$ and $w_1 \succ_2 w_2$.*

The minimax merger can be defined as follows.

Proposition 3. *Let $(E_1, \cdots, E_n)$ and $(E'_1, \cdots, E'_m)$ be the ordered partitions of $\succeq_d$ and $\succeq_b$ respectively. The result of merging $\succeq_d$ and $\succeq_b$ is $\succeq = (E''_1, \cdots, E''_{n\times m})$ such that if $\omega \in E_i$ and $\omega \in E'_j$ then $\omega \in E''_{(i-1)*m+j}$.*

Example 4. (continued) The minimax merger of the norm specification is $\{\{mp\}, \{\neg mp\}, \{m\neg p\}, \{\neg m\neg p\}\}$.

The principle of this merger is similar to minimax merger. The dictator here is the pre-order associated to the 'pessimistic' preference specification and computed following the maximal specificity principle.

Definition 11. $w_1 \succ w_2$ *iff* $w_1 \succ_2 w_2$ *or* $w_1 \sim_2 w_2$ *and* $w_1 \succ_1 w_2$.

Example 5. Consider the 'optimistic' $O(p)$ and the 'pessimistic' $O(m)$. The merger is $\{\{mp\}, \{m\neg p\}, \{\neg mp\}, \{\neg m\neg p\}\}$.

The problem of handling preferences on controllable variables and uncontrollable variables separately is that it is not possible to express *interaction* between the two kinds of variables. For example my decision on whether I will work hard to finish a paper (which is a controllable variable) depends on the uncontrollable variable "money", decided by my boss. If my boss accepts to pay the conference fees then I have to work hard to finish the paper. We therefore consider in the remainder of this paper preference formulas with controllable and uncontrollable variables.

A general approach would be to define 'optimistic' and 'pessimistic' preference specifications on any combination of controllables and uncontrollables, such as an 'optimistic' preference $O^-(p \wedge x)$ or even $O^+(p)$. However, this approach blurs the idea that 'optimistic' reasoning is restricted to controllables, and 'pessimistic' reasoning is restricted to uncontrollables. Mixed 'optimistic' and 'pessimistic' norms are defined as follows.

Definition 12 (Mixed norm specification). *A conditional 'optimistic' obligation is a formula of the form* $\{O(x_i|p_i \wedge y_i) \mid i = 1, \cdots, n, p_i \in \mathcal{L}_\mathcal{U}, x_i, y_i \in \mathcal{L}_\mathcal{C}\}$.

We merge the two pre-orders using the symmetric merger operator since there is no reason to give priority either to $\succeq_d$ or to $\succeq_b$.

6 Related work

6.1 Permission

One may wonder whether there is a role for permissive norms in deontic logic, once we have obligations. The issue is whether there is a need to distinguish between the absence of a prohibition and an explicit permission. Several researchers have doubted whether there is such a need, which may explain why the study of permission has gotten so little attention in the deontic logic literature, despite the fact that deontic logic started by von Wright's observation that the relation between obligation and permission is analogous to the relation between necessity and possibility, and despite the fact that the earliest papers used permission instead of its dual obligation.

The distinction between weak and strong permission is well known in deontic logic, and the permission in DSDL3 is a standard example of a weak permission. Basically, the argument is that a weak permission is only the absence of an obligation, and this is precisely how permissions are defined in DSDL3. Moreover, we have as a theorem $O(p|q) \vee P(\neg p|q)$, indicating that every proposition is normed; a typical property of weak permission. However, it is a priori less clear whether these arguments still hold in non-monotonic DSDL3. We therefore consider two more detailed discussions on the distinction between weak and strong permissions.

However, there has been a convincing argument that permissions are a distinct kind of norms, besides obligations. Bulygin [6] observes that in a setting with higher and lower authorities, a higher authority needs to issue strong permissions to delimit the power of the lower authorities. So when there are two agents only and we consider DSDL3 as a compact specification language used by agent 2 (an authority) to define a pre-order for agent 1 (his ideal and sub-ideal states), then we only have to use obligations. However, as Bulygin argues, the picture is completely different when we consider three agents, for example in hierarchical normative systems, where a higher authority agent 0 limits the pre-orders agent 1 can prescribe to agent 2. Agent 0 can say now, for example, that agent 1 is not allowed to oblige p for agent 2. Bulygin's game among three agents is challenging for the development of deontic logic for normative multi-agent systems, but its implications for non-monotonic DSDL3 are not clear to us at this moment.

In Makinson and van der Torre's analysis of permission in their input/output logic framework, the distinction between weak and strong permissions has been made explicit (because the set of norms in the input/output logic framework is explicit). Maybe the norm specification used in non-monotonic DSDL3 may also be seen as such an explicitly represented set of norms.

6.2 Non-monotonic logic

Our algorithm generalizes the algorithm of Benferhat *et al.* [2] which captures "equal preferences" , denoted $p = q$, which stands for "all best p worlds are q worlds and all best q worlds are p worlds". These equivalences can be represented in our framework by two non-strict preferences $p \geq q$ and $q \geq p$, but our non-strict preferences cannot be represented in their framework.

7 Summary

We study specificity principles for non-monotonic extensions of Bengt Hansson's standard dyadic deontic logic 3, known as DSDL3, with both controllable and uncontrollable propositions. This extension is important in artificial normative systems, maybe more than in alternative applications of DSDL3 in default reasoning, qualitative decision or preference logic. Permissions play an important role in normative multi-agent systems, for example when higher and lower authorities are distinguished in normative multi-agent systems. Moreover, the distinction between ought-to-be and ought-to-do is central in agent theory too.

We introduce an efficient algorithm for minimal specificity which not only covers obligations but also permissions. The extension with permissions is more complicated, because we cannot directly define the equivalence classes of the pre-order, but we need a second loop in the algorithm to deal with the permissions. The algorithm may be further extended with other kinds of obligations and permissions, for example with ceteris paribus obligations (p is obliged if each p world is preferred to each $\neg p$ world, ceteris paribus).

Moreover, we introduce ways to combine ought-to-be and ought-to-do obligations in DSDL3 extended with the distinction between controllable and uncontrollable propositions. We illustrate our approach for algorithms for minimal and maximal specificity for DSDL3 with controllable and uncontrollable propositions, based on 'optimistic' and 'pessimistic' reasoning respectively. Alternative ways to combine ought-to-be and ought-to-do obligations in non-monotonic DSDL3 are subject of further research.

An assumption of this paper has been that DSDL3 is established as a deontic logic, and that its non-monotonic mechanisms are needed to deal with either violations or exceptions. However, a referee has suggested to us that in a deontic logic it does not seem to make sense to assume that worlds are as good as possible and that agents will see to it that the best (or worst) worlds are realized. Moreover, he or she observes that these criticisms seem to only apply to deontic uses of DSDL3. It may thus be that we have not sufficiently separated applications of DSDL3 to deontic reasoning from applications to default reasoning and decision making. We have left this issue for further research.

References

1. S. Benferhat, D. Dubois, and H. Prade. Possibilistic and standard probabilistic semantics of conditional knowledge bases. *Logic and Computation*, 9:6:873–895, 1999.
2. S. Benferhat, D. Dubois, and H. Prade. Towards a possibilistic logic handling of preferences. *Applied Intelligence*, 14(3):303–317, 2001.
3. R. Booth and J. B. Paris. A note on the rational closure of knowledge bases with both positive and negative knowledge. *Journal of Logic, Language and Information*, 7(2), 1998.
4. C. Boutilier. Conditional logics of normality : a modal approach. *Artificial Intelligence*, 68:87–154, 1994.
5. C. Boutilier. Toward a logic for qualitative decision theory. In *Proceedings of the 4th International Conference on Principles of Knowledge Representation, (KR'94)*, pages 75–86, 1994.
6. E. Bulygin. Permissive norms and normative systems. In A. Martino and F. S. Natali, editors, *Automated Analysis of Legal Texts*, pages 211–218. Publishing Company, Amsterdam, 1986.
7. L. Cholvy and C. Garion. Deriving individual obligations from collective obligations. In *Procs of AAMAS 2003*, pages 962–963, 2003.
8. L. Cholvy and C. Garion. Desires, norms and constraints. In *Procs of AAMAS 2004*, pages 724–731, 2004.
9. Bengt Hansson. An analysis of some deontic logics. *Nos*, 3:373–398, 1969. Reprinted in Hilpinen (1971), pages 121 -147.
10. S. Kaci and L. van der Torre. Algorithms for a nonmonotonic logic of preferences. In *Procs of 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, LNCS 3571, pages 281–292. Springer, 2005.
11. S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning , preferential models and cumulative logics. *Artificial Intelligence*, 44 (1):167– 207, 1990.
12. J. Lang, L. Van Der Torre, and E. Weydert. Utilitarian desires. *Autonomous Agents and Multi-Agent Systems*, 5:329–363, 2002.

13. D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55(1):1–60, 1992.
14. D. Lewis. *Counterfactuals*. Blackwell, 1973.
15. J. Pearl. System Z: A natural ordering of defaults with tractable applications to default reasoning. In R. Parikh. Eds, editor, *Proceedings of the 3rd Conference on Theoretical Aspects of Reasoning about Knowledge (TARK'90)*, pages 121–135. Morgan Kaufmann, 1990.
16. H. Prakken and M.J. Sergot. Dyadic deontic logic and contrary-to-duty obligations. In D. Nute, editor, *Defeasible Deontic Logic*, volume 263 of *Synthese Library*, pages 223–262. Kluwer, 1997.
17. Y. Shoham. Nonmonotonic logics: Meaning and utility. In *Procs of IJCAI 1987*, pages 388–393, 1987.
18. Wolfgang Spohn. An analysis of Hanssons dyadic deontic logic. *Journal of Philosophical Logic*, 4:237 – 252, 1975.
19. W. van der Hoek and M. Wooldridge. On the logic of cooperation and propositional control. *Artif. Intell.*, 164(1-2):81–119, 2005.
20. L. van der Torre and Y. Tan. Contrary-to-duty reasoning with preference-based dyadic obligations. *Annals of Mathematics and Artificial Intelligence*, 27:49–78, 1999.