Artificial Intelligence and Law **7:** 51–67, 1999.

Diagnosis and decision making in normative reasoning

LEENDERT W.N. VAN DER TORRE
Max Planck Institute for Computer Science, Im Stadtwald, D-66123 Saarbrücken
E-mail: torre@mpi-sb.mpg.de

YAO-HUA TAN
EURIDIS, *Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam*
E-mail: ytan@fac.fbk.eur.nl

Abstract. Diagnosis theory reasons about incomplete knowledge and only considers the past. It distinguishes between violations and non-violations. Qualitative decision theory reasons about decision variables and considers the future. It distinguishes between fulfilled goals and unfulfilled goals. In this paper we formalize normative diagnoses and decisions in the special purpose formalism DIO(DE)^2 as well as in extensions of the preference-based deontic logic PDL. The DIagnostic and DEcision-theoretic framework for DEontic reasoning DIO(DE)^2 formalizes reasoning about violations and fulfillments, and is used to characterize the distinction between normative diagnosis theory and (qualitative) decision theory. The extension of the preference-based deontic logic PDL shows how normative diagnostic and decision-theoretic reasoning – i.e. reasoning about violations and fulfillments – can be formalized as an extension of deontic reasoning.

1. Introduction

In the AI and Law literature it is discussed whether deontic logic should be used to formalize legal reasoning (and normative reasoning in general). Jones and Sergot (Jones and Sergot, 1992, 1993) argue that deontic logic is a useful knowledge representation language when the modeler wants to formalize reasoning about violations and obligations that arise as a result of these violations, the so-called contrary-to-duty obligations. McCarty (1994) observes that 'one of the main features of deontic logic is the fact that actors do not always obey the law. Indeed, it is precisely when a forbidden act occurs, or an obligatory action does not occur, that we need the machinery of deontic logic, to detect a violation and to take appropriate action.' These claims are not undisputed. For example, Bench-Capon (1994) argues that in many cases, including the widely discussed Imperial College Library Regulations, the representation of regulations as norms is 'at best unhelpful and at worst misleading.' In our opinion, this discussion on the use of deontic logic to formalize legal reasoning should be extended to cover other theories of normative reasoning (von

Wright, 1983). To this end, we discuss two formalizations of normative diagnosis and decision theory.

1. We use the special purpose formalism DIO(DE)^2 to formalize the distinction between normative diagnosis and decision theory.* Advantages of this formalization over the following one is that conceptually the distinction between diagnosis and decision theory is much more clear and explicit, and it is computationally more efficient, because logical relations between norms do not have to be taken into account.
2. We use the preference-based deontic logic PDL with additional principles to show that normative diagnosis and decision theory can be formalized as extensions of a suitable deontic logic. Advantages of this formalization over the previous one is its ability to be embedded in more general forms of normative reasoning such as for example normative conflict detection and resolution.

In this paper we use the DIagnostic and DEcision-theoretic framework for DEontic reasoning DIO(DE)^2 to discuss the distinction between diagnostic reasoning and decision-theoretic reasoning, represented in Figure 1. It illustrates that the two

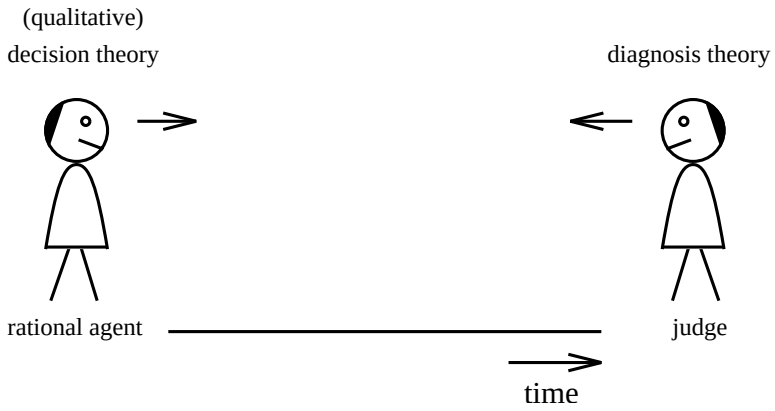

Figure 1. Reasoning with norms.

theories have different temporal perspectives. Diagnosis theory reasons about violations, and checks systems against given principles. In particular, it reasons about the past with incomplete knowledge (if everything is known, then a diagnosis is completely known). Diagnosis theory formalizes the hypothetical as-if reasoning of a judge or public prosecutor when she checks legal systems against legal principles. Qualitative decision theory describes how norms influence behavior and is based on the concept of agent rationality. In contrast to diagnostic theories, a (qualitative) decision theory reasons about the future. The main characteristic of

* See (van der Torre et al., 1997; van der Torre and Tan, 1997a) for the subtle distinctions between the present framework and its predecessors, where we discuss ideas we previously proposed as well as observations of other researchers on this work, and we show how these are incorporated in the new version of the framework.

qualitative decision theory is that it is goal oriented reasoning, for example in planning. This reasoning is based on the application of strategies, which can be considered as qualitative versions of the 'maximum utility' criterion. DIO(DE)^2 is built from first principles and contains two main ingredients. First, it contains representations of violations to formalize the reasoning about violations of the DIagnostic framework for DEontic reasoning DIODE (Tan & van der Torre, 1994a, 1994c), Reiter's theory of diagnosis (Reiter, 1987) applied to normative systems. Second, it contains representations of fulfilled obligations, which make it possible to formalize reasoning about goals.

Moreover, in this paper we use the preference-based deontic logic PDL to show how deontic logic can be used as a component in normative diagnosis theory as well as qualitative decision theory.[⋆] Normative reasoning is more than deontic logic, because deontic logic only tells which obligations can be derived from a set of other obligations. In particular, it characterizes the logical relations between obligations. For example, in most deontic logics the conjunction $p \wedge q$ is obliged, if both p and q are obliged. Logical relations between obligations can be used in a formalism that explains the effect of norms on behavior. Diagnostic reasoning and decision-theoretic reasoning is formalized by adding assumptions to PDL. For example, if you approach a green traffic light on a square, and another car approaches from the left where the light is red, then you assume that the other car will stop. This assumption cannot be explained by a deontic logic. McCarty (1994) observes that for purposes of planning, it is often useful to assume that actors *do* obey the law. He calls this the *causal assumption*,[⋆⋆] since it enables us to predict the actions that *will* occur by reasoning about the actions that *ought* to occur. McCarty concludes that if we adopt the causal assumption, we can use the machinery of deontic logic to reason about the physical world. We show that the formalization of normative diagnoses and decisions in PDL is closely related to their formalization in DIO(DE)^2.[‡]

[⋆] The diagnosis of a normative system can use a formalism to represent norms and additional assumptions or principles to do the diagnosis. Similarly, normative planning can use a special formalism to represent norms and additional principles to do the planning. For example, Reiter's diagnosis (Reiter, 1987) is basically a minimization principle (called the principle of parsimony). Similarly, qualitative decision theory has a formalism for representing norms (or goals) and additional assumptions or principles to reason with them. This raises several interesting questions, which are not addressed in this paper. Is such a special purpose formalism a deontic logic? How do they stand the test against the Chisholm paradox, the paradox of the gentle murderer, the problem of how to represent permissions, the problem of conflicting obligations? What are the structural similarities and distinctions between the different formalisms?

[⋆⋆] For example, many of the norms of social morality 'codify' behavior regularities that are spontaneously wanted by most members of the group. The norms sanction the 'abnormal' behavior of the minority. Legal norms may be a further codification of such norms of social morality. They may also be norms protecting the interest of certain groups, who again may have a spontaneous wish to abide to these norms.

[‡] Qualitative decision theory is based upon the concept of (internal) preference. This preference is a kind of desire, i.e. it is an endogenously motivating mechanism (coming from the agent itself).

DIO(DE)^2 and PDL with additional assumptions are useful for applications, because they can be used as the basis of a knowledge representation language in a decision support system. However, the formalisms do not model the legal practice like, for example, argumentation theory (Hage, 1996; Prakken & Sartor, 1996) or debating systems (Gordon, 1995), because the formalisms do not formalize the legal context. For example, one feature of actual normative diagnosis in law is that it is usually a top-down process instead of a consistency check. The legal context usually provides just a few (often even just one) potential diagnoses. In criminal law the prosecutor's indictment leaves just a few options open, while in civil proceedings it is the parties who, through their claims, focus the debate on just a few issues. To describe this type of reasoning it seems sufficient to have a module that, *given* an obligation 'p ought to be the case,' checks whether $\neg p$ can be derived. The theory provided by DIO(DE)^2 seems overkill here. However, the prosecutor may use a decision support system based on DIO(DE)^2 to select an indictment, and in civil proceedings the parties may use a decision support system to select the claims on which they thereafter focus the debate. For these decision problems, the formalisms discussed in this paper are useful for applications.

2. DI**O(**DE**)**^{2}

In this section we discuss the DIagnostic and DEcision-theoretic framework for DEontic reasoning DIO(DE)^2 and we illustrate it by several examples.

2.1. DIAGNOSIS THEORY

We first discuss diagnosis theory and how this theory can be used to formalize normative reasoning. The model-based reasoning approach to diagnosis has been studied for several years. Numerous applications have been built, most of all for diagnosis of physical devices. The basic paradigm is the interaction of prediction and observation. Predictions are expected outputs given the assumption that all the components are working properly. If a discrepancy between the output of the system (given a particular input) and the prediction is found, then the diagnosis procedure will search for defects in the components of the system. The DIagnostic framework for DEontic reasoning DIODE introduced in (Tan & van der Torre, 1994a, 1994c) formalizes deontic reasoning as a kind of diagnostic reasoning. Notice that DIODE is not a deontic logic (it does not describe which obligations follow from a set of obligations) and it should not be considered as such. On the

Therefore, it is not a natural candidate for dealing with normative decision-making, since a norm is by definition exogenous, in the sense that it is something the agent would not spontaneously want (Lang, 1996). This again raises interesting questions, which are not discussed in this paper. How do agents work out norms in terms of gains and losses? What are the gains of observing norms? How do they learn the effects of norms and how do they reason about these effects? Which rules are implied, which ingredients enable agents to make normative decisions? In which way does a normative decision maker differ from an ordinary decision maker, if any?

other hand, since diagnosis reasons about violations and deontic logic is useful to model situations where violations are important (Jones & Sergot, 1992), it makes sense to have a deontic framework for diagnosis like DIODE. The framework treats norms as components of a system to be diagnosed; hence the system description becomes a norms description.

The contribution of Reiter to diagnosis theory is widely accepted. His *consistency-based approach* (Reiter, 1987) is the first one to model the model-based reasoning approach to diagnosis. The main goal is to eliminate system inconsistency by identifying the minimal set of abnormal components that is responsible for the inconsistency, which are represented by the abnormality predicate Ab. That is, reasoning about diagnoses is based on the following assumption of diagnostic reasoning.

> **Principle of parsimony** is the conjecture that the set of faulty components is minimal (with respect to set inclusion).

Related to a diagnosis is a set of measurements, the predictions given the assumption that most components are working properly.

DEFINITION 1 (Diagnosis). A *system* is a pair (COMP, SD) where COMP, the *system components*, is a finite set of constants denoting the components of the system, and SD, the *system description*, is a set of first-order sentences. An *observation* of a system is a finite set of first-order sentences. A system to be diagnosed, written as (COMP, SD, OBS), is a system (COMP, SD) with observation OBS. A *diagnosis* for (COMP, SD, OBS) is a minimal (with respect to set inclusion) set $\Delta \subseteq \text{COMP}$ such that

$$\text{CONTEXT}_\Delta = \text{SD} \cup \text{OBS} \cup \{Ab(c) \mid c \in \Delta\} \cup \{\neg Ab(c) \mid c \in \text{COMP} - \Delta\}$$

is consistent. A diagnosis Δ for (COMP, SD, OBS) *predicts* a *measurement* Π if and only if $\text{CONTEXT}_\Delta \models \Pi$.

We refer to the base logic of DIODE as $\mathcal{L}_V$, and the fragment of $\mathcal{L}_V$ without violation constants as $\mathcal{L}$. We write $\models$ for entailment in $\mathcal{L}_V$. The definition of minimal violated-norm set is analogous to the definition of diagnosis. Just as we can have multiple diagnoses with respect to the same (COMP, SD, OBS), we can have multiple minimal violated-norm sets Δ with respect to (NORMS, ND, FACTS). We can have more than one minimal violation state, which reflects that we can have different situations that are optimal, i.e. as ideal as possible. Ramos and Fiadeiro (1996) observe that in normative diagnostic reasoning Reiter's theory of diagnosis focuses on the *minimal* sets of violations. They argue that the underlying assumption 'innocent until proven guilty' is not always the right one. For example, if constraints of a management process are modeled as obligations that have to be fulfilled, then the assumption 'guilty until proven innocent' might be more reasonable. In criminal proceedings, the defendant of the accused might want

to argue for all possible violations of procedural norms by the prosecution. They therefore distinguish between potential diagnoses (any subset of NORMS such that CONTEXT_Δ is consistent) and minimal and maximal diagnoses (or violated-norm sets, called benevolent and exigent diagnoses by Ramos & Fiadeiro).

DEFINITION 2. (DIODE). A *normative system* is a tuple NS = (NORMS, ND) with NORMS, a finite set of constants $\{n_1, \ldots, n_k\}$ denoting *norms*, and ND, the *norms description*, a set of first-order $\mathcal{L}_V$ sentences $\neg V(n_i) \leftrightarrow (\beta \rightarrow \alpha)$ denoting *obligations*. A *normative system to be diagnosed* is a tuple NSD = (NORMS, ND, FACTS) with NS = (NORMS, ND), a normative system, and FACTS, a set of first-order $\mathcal{L}$ sentences that describe the facts. A *potential diagnosis* Δ of NSD is a subset of NORMS such that

$$\text{CONTEXT}_\Delta = \text{ND} \cup \text{FACTS} \cup \{V(n_i) \mid n_i \in \Delta\} \cup \{\neg V(n_i) \mid n_i \in \text{NORMS} - \Delta\}$$

is consistent. A *minimal (maximal) diagnosis* Δ of NSD is a minimal (maximal) (with respect to set inclusion) subset of NORMS such that CONTEXT_Δ is consistent. The set of *contextual obligations* of a minimal diagnosis Δ of a normative system to be diagnosed NSD is $\text{CO}_\Delta = \{\alpha \mid \alpha \in \mathcal{L}, \text{CONTEXT}_\Delta \models \alpha\}$.

Again we emphasize that DIODE does not formalize logical relations between norms, and thus is not a deontic logic. In (van der Torre & Tan, 1997a) we discuss the relation between DIODE and Anderson's reduction of Standard Deontic Logic to alethic modal logic (Anderson, 1958). Moreover, we discuss the relation between diagnostic reasoning and deontic logic later in this paper. Two aspects of the diagnostic approach are explicitly distinguished in DIODE. The first aspect concerns violation detection and looking backward perspective. The second aspect is the principle of parsimony, a reasoning strategy to deal with incomplete information in violation detection. This principle is formalized by the minimality condition, and formalizes the 'innocent until proven guilty' assumption. This difference between two aspects might correspond to the judge view and the lawyer view on a normative system. In this perspective, the judge only checks whether norms are violated, and it is the lawyer that argues for a *minimal* set of violations (arguing that the burden of proof is with the prosecution). The following example adapted from (Smith, 1994) illustrates DIODE.

EXAMPLE 1 (Convention on contracts). Consider the following normative system 'party A should deliver in time' (d), 'if the party does deliver in time, then it should not give notice' ($\neg n$), 'if the party does not deliver then it should give notice' (n) and the party does not deliver in time ($\neg d$).

- NORMS = $\{n_1, n_2, n_3\}$,
- ND = $\{(\neg V(n_1) \leftrightarrow d), (\neg V(n_2) \leftrightarrow (d \rightarrow \neg n)), (\neg V(n_3) \leftrightarrow (\neg d \rightarrow n))\}$,
- FACTS = $\{\neg d\}$.

It is easily checked that CONTEXT_Δ implies $\{\neg d, V(n_1), \neg V(n_2), \neg V(n_3) \leftrightarrow n\}$. There are two potential diagnosis, $\Delta_1 = \{n_1\}$ and $\Delta_2 = \{n_1, n_3\}$, where Δ_1 is the minimal violated-norm set and Δ_2 the maximal one. The context of Δ_1 implies n and the context of Δ_2 implies $\neg n$. Hence, the minimal violated-norm set Δ_1 implies that it is assumed that the party gives notice (innocent: the third norm is not violated) and the maximal violated-norm set Δ_2 assumes that it does not (guilty: the third norm is violated).

2.2. QUALITATIVE DECISION THEORY

In the usual approaches to planning in AI, a planning agent like a robot is provided with a description of some state of affairs, a *goal state*, and charged with the task of discovering (or performing) some sequence of actions to achieve that goal. *Context-sensitive* goals are used to formalize objectives faced by the robot which reflect graded criteria, such as time taken to fill a tank or amount of fluid spilled. In realistic planning situations the robot's objectives can be satisfied to varying degrees, and an agent will frequently encounter goals that it cannot achieve. Context-sensitive goals are formalized with basic concepts provided by decision theory (Dean & Wellman, 1991; Doyle & Wellman, 1991; Boutilier, 1994).

Reasoning about goals is formalized in DIO(DE)^2, the DIagnostic and DEcision-theoretic framework for DEontic reasoning that extends DIODE.* The crucial aspect of goals formalized in DIODE is that goals can be fulfilled. The formal language of DIO(DE)^2 contains a fulfilled-norm predicate (F) to formalize these fulfilled goals. The norm 'α should be (done) if β is (done)' is formalized by $\neg V(n) \leftrightarrow (\beta \rightarrow \alpha)$ and $F(n) \leftrightarrow (\beta \wedge \alpha)$. The formalization shows that fulfillment of a context-sensitive goal is different from non-violation of such a goal. As a consequence, knowledge referring to fulfilled goals cannot be expressed in DIODE. A theory of diagnosis like DIO(DE)^2 is based on the distinction between violated and non-violated, whereas a (qualitative) decision theory is based on the distinction between fulfilled and non-fulfilled. DIO(DE)^2 combines reasoning about violated and fulfilled norms. Hence, it combines reasoning about the past (violated versus non-violated) with reasoning about the future (already fulfilled versus not yet fulfilled). As illustrated in Figure 1, DIO(DE)^2 combines the diagnostic reasoning of a judge with the planning reasoning of a rational agent. For a comparison between DIO(DE)^2 and Ramos and Fiadeiro's DDD (Ramos & Fiadeiro, 1996, 1998) see (van der Torre et al., 1997).

* Goals serve a dual role in most planning systems, capturing aspects of both *intentions* and *desires* (Doyle, 1980). Besides expressing the desirability of a state, adopting a goal represents some commitment to pursuing that state. For example, accepting a proposition as an achievement task commits the agent to finding some way to accomplish this objective, even if this requires adopting some subtasks that may not corresponds to desirable propositions themselves (Dean & Wellman, 1991). In our semantical interpretation, we concentrate exclusively on the role of expressing desirability, recognizing that the result is only a partial account of the use of goals in planning systems.

In the following Definition 3 there are two orderings on pairs of norms. The first ordering $\sqsubseteq$ gives the potential diagnoses by determining the active norms, i.e. the norms which are in force (which are the minimal elements in the ordering $\sqsubseteq$). The second ordering $\leq$ gives the minimal and maximal diagnoses by comparing the pairs of norms in a similar way as diagnoses are compared in DIODE.

DEFINITION 3 (DIO(DE)^2). A *normative system* is a tuple NS = (NORMS, ND_F) where ND_F, the *norms description*, is a set of *conditional obligations*

$$(\neg V(n_i) \leftrightarrow (\beta \rightarrow \alpha)) \wedge (F(n) \leftrightarrow (\beta \wedge \alpha))$$

Let NSD = (NORMS, ND_F, FACTS) be a normative system to be diagnosed. A *fulfilled-violated set* (Δ_f, Δ_v) of NSD is a pair of subsets of NORMS such that

$$\begin{aligned}\text{CONTEXT}_\Delta = \ & \text{ND}_F \cup \text{FACTS} \\ & \cup \{F(n_i) \mid n_i \in \Delta_f\} \cup \{\neg F(n_i) \mid n_i \in \text{NORMS} - \Delta_f\} \\ & \cup \{V(n_i) \mid n_i \in \Delta_v\} \cup \{\neg V(n_i) \mid n_i \in \text{NORMS} - \Delta_v\}\end{aligned}$$

is consistent. Let $\sqsubseteq$ be the ordering on fulfilled-violated sets defined by the relation $(\Delta_f, \Delta_v) \sqsubseteq (\Delta'_f, \Delta'_v)$ if and only if $\Delta_f \subseteq \Delta'_f$ and $\Delta_v \subseteq \Delta'_v$. A *potential diagnosis* (Δ_f, Δ_v) of NSD is a fulfilled-violated set that is minimal in the ordering $\sqsubseteq$. Let $\leq$ be the ordering on potential diagnoses such that $(\Delta_f, \Delta_v) \leq (\Delta'_f, \Delta'_v)$ if and only if $\Delta'_f \subseteq \Delta_f$ and $\Delta_v \subseteq \Delta'_v$. A *minimal (maximal) diagnosis* (Δ_f, Δ_v) of NSD is a potential diagnosis that is minimal (maximal) in the ordering $\leq$.

The following example is adapted from Example 1 and illustrates DIO(DE)^2.

EXAMPLE 2 (Transitivity). Consider the following normative system of the two obligations 'party A has to deliver the goods (d)' and ' if party A delivers the goods, then it has to give notice of the expected arrival date to party B in advance (n),' together with the fact that party A did not give notice.

- NORMS = $\{n_1, n_2\}$,
- $\text{ND}_F = \left\{ \begin{array}{l} (\neg V(n_1) \leftrightarrow d) \wedge (F(n_1) \leftrightarrow d), \\ (\neg V(n_2) \leftrightarrow (d \rightarrow n)) \wedge (F(n_2) \leftrightarrow (d \wedge n)) \end{array} \right\}$,
- FACTS = $\{\neg n\}$.

The potential diagnoses are $(\Delta_f, \Delta_v) = (\emptyset, \{n_1\})$ and $(\Delta_f, \Delta_v) = (\{n_1\}, \{n_2\})$, and their contexts imply respectively $\{\neg d, \neg n\}$ and $\{d, \neg n\}$. Both potential diagnoses are minimal.

In the first paragraph of this section we already observed that context-sensitive goals have utilitarian or preference-based (i.e. decision-theoretic) semantics to formalize different degrees of goal violation (Dean & Wellman, 1991; Doyle & Wellman, 1991; Boutilier, 1994). The obvious formalization of different degrees is to

introduce a set of violation predicates, one for each degree of violation. Consider for example the libel article, that discriminates between two types of violation, insults in private and insults in public, see also (Tan and van der Torre, 1994b). If we ignore the exception clause, then the libel article can be formalized in an extension of DIO(DE)^2 by the normative system of the obligation 'it is forbidden to insult' and two ways to violate it: in private (r) and in public ($\neg r$).

- NORMS = $\{n\}$, and
- $\text{ND}_F = \{(\neg V_1(n) \leftrightarrow (i \rightarrow \neg r)) \wedge (\neg V_2(n) \leftrightarrow (i \rightarrow r)) \wedge (F(n) \leftrightarrow \neg i)\}$.

However, this extension of DIO(DE)^2 introduces additional complexity. The following example illustrates how we can formalize the libel article in DIO(DE)^2 by introducing sub-norms for every way in which a norm can be violated.

EXAMPLE 3 (Degrees). Consider the following normative system of DIO(DE)^2.

- NORMS = $\{n_1, n_2\}$,
- $\text{ND}_F = \left\{ \begin{array}{l} (\neg V(n_1) \leftrightarrow \neg i) \wedge (F(n_1) \leftrightarrow \neg i), \\ (\neg V(n_2) \leftrightarrow (i \rightarrow r)) \wedge (F(n_2) \leftrightarrow (i \wedge r)) \end{array} \right\}$,
- FACTS = $\{i\}$.

The potential diagnosis are $(\Delta_f, \Delta_v) = (\{n_2\}, \{n_1\})$ and $(\Delta_f, \Delta_v) = (\emptyset, \{n_1, n_2\})$, where the first is minimal and the latter maximal. The context of the minimal diagnosis implies that the insult occurred in private (r), and the context of the maximal one implies that the insult occurred in public ($\neg r$).

The previous example illustrates that DIO(DE)^2 can also formalize normative diagnosis when the legal code refers to a non-deontic category (libel), i.e. a category that does not contain deontic terms, by making explicit when behavior is classified in this category (insults). Other examples of such non-deontic categories are felony, misdemeanor, tort, and insufficient care. Prohibitions are often stated as such non-deontic categories, to which certain legal consequences are attached (e.g. punishability in criminal law, liability in civil law). If DIO(DE)^2 is to be applied to regulations that contain such concepts, then they all have to be (at least partially) characterized by observable behavior (in other words, they have to be translated into deontic terms). We leave it an open question whether this is feasible in practice. Obviously, it will be much harder to characterize a case of insufficient care than characterizing a case of libel as in Example 3, which is itself already a simplification of the legal practice (because not all insults are a case of libel).

3. Normative Diagnostic and Decision-Theoretic Reasoning as Extensions of Deontic Reasoning

Ramos and Fiadeiro (1996, 1998; van der Torre et al., 1997) show that the diagnostic reasoning can also be formalized as an extension of deontic reasoning by adding principles to a suitable deontic logic. Deontic logic characterizes logical relations between obligations. However, it does not explain how norms affect the behavior of rational agents. From Op you cannot infer whether somebody will actually perform p. This is no critique on deontic logic, it is just an observation. Deontic logic was never intended to explain this effect of norms on behavior. However, if we want to explain all the different aspects of normative reasoning, then we need more formalisms than just deontic logic (von Wright, 1983). This observation is commonly accepted, and in AI and Law logic has been embedded (explicitly or implicitly) in e.g. models of analogical reasoning, defeasible argumentation, procedural models for dispute, spotting issues and heuristics for dispute. Moreover, it is also in general AI widely accepted that reasoning is more than deductive logic.

The following formalization in our preference-based deontic logic PDL illustrates that normative diagnosis theory as well as qualitative decision theory can be viewed as extensions of deontic logic. This formalization illustrates how logical relations between obligations can be used in a formalism that explains the effect of norms on behavior. In both cases the formalism contains extra principles that are added to a deontic logic basis. For example, in the case of diagnosis theory one of the principles that can be added to deontic logic is the parsimony principle, i.e. the assumption that as few obligations as possible are violated. There is nothing paradoxical in the claim that on the one hand these formalisms explain aspects of normative behavior that deontic logic does not, whereas deontic logic is still an essential component of these theories. In the same sense physics can explain phenomena that mathematics cannot, whereas mathematics is still an essential component of physics. Note that the two formalisms DIODE and DIO(DE)^2 are not based on deontic logic, which shows that the normative diagnosis and decision theory do not have to be formalized as extensions of deontic logic.

3.1. PREFERENCE-BASED DEONTIC LOGIC

In the modal preference semantics of PDL, the accessibility relation is interpreted as a preference relation. For example, $w_1 \leq w_2$ has to be read as 'world w_1 is at least as good as world w_2.' It is a well-known problem from preference logics that we cannot define an obligation Op as a strict preference of p over $\neg p$, because two obligations Op_1 and Op_2 would conflict for $p_1 \wedge \neg p_2$ and $\neg p_1 \wedge p_2$. According to the first obligation Op_1, worlds satisfying $p_1 \wedge \neg p_2$ are preferred to worlds satisfying $\neg p_1 \wedge p_2$, and according to the obligation Op_2 vice versa. The two preference statements are contradictory. This motivates the following weaker definition: an obligation p is the absence of a preference of $\neg p$ over p, see (Tan & van der Torre, 1996; van der Torre & Tan, 1997b).

DEFINITION 4 (Preference-based obligations). A Kripke model $M = \langle W, \leq, V \rangle$ consists of W, a set of worlds, $\leq$ a binary reflexive accessibility relation interpreted as a preference relation, and V, a valuation of the propositions at the worlds. We have $M \models O(\alpha \mid \beta)$ if and only if

1. for all worlds w and w' such that $M, w \models \alpha \wedge \beta$ and $M, w' \models \neg\alpha \wedge \beta$, we have $w' \not\leq w$, and
2. there are such worlds w and w'.

In the preference-based logic of obligations in the previous Definition 4, conflicts like $Op \wedge O\neg p$ are consistent. In (Tan & van der Torre, 1996; van der Torre & Tan, 1997b) we discuss how to use other operators which make conflicts inconsistent. For consistent sets of premises, these operators induce the same preference orderings as the operators discussed in this paper, and for the purposes of this paper the distinction is irrelevant. The following definition illustrates that the modal logic PDL can be used as the basis of a normative diagnosis or decision theory. The definition consists of two steps. First, the preference-based deontic logic is used to determine the active obligation set.⋆ Second, the active obligations are used to define minimal and maximal diagnoses.⋆⋆

DEFINITION 5 (Normative diagnosis). An obligation system to be diagnosed is a tuple OSD = (OBL, FACTS) with:

1. OBL, a finite set of modal sentences denoting conditional obligations $O(\alpha|\beta)$,
2. FACTS, a finite set of propositional sentences.

The *active obligation set* AO is the set of obligations:

$$\text{AO} = \{O(\alpha|\beta) \mid \text{OBL} \cup \text{FACTS} \models_{\text{PDL}} O(\alpha|\beta) \wedge \beta\}$$

A *minimal active obligation set* MAO is a subset of AO such that $\text{MAO} \models_{\text{PDL}} \text{AO}$ and for all subsets S of MAO we have $S \not\models_{\text{PDL}} \text{AO}$. A potential (minimal/maximal) diagnosis Δ is a (minimal/maximal) subset of a minimal active obligation set MAO such that

$$\text{CONTEXT}_\Delta = \text{OBL} \cup \text{FACTS} \cup \{\neg\alpha \mid O(\alpha \mid \beta) \in \Delta\} \cup \{\alpha \mid O(\alpha \mid \beta) \in \text{MAO} - \Delta\}$$

is consistent.

The set $\{\neg\alpha \mid O(\alpha \mid \beta) \in \Delta\}$ of CONTEXT_Δ is related to the set of violated norms Δ_v in DIO(DE)^2 and the set $\{\alpha \mid O(\alpha \mid \beta) \in \text{MAO} - \Delta\}$ is related to

⋆ The active obligation set can also be defined in the language of the deontic logic PDL. In particular, we can use the factual detachment derivation $\beta \wedge O(\alpha \mid \beta) \rightarrow O_a(\alpha \mid \beta)$.

⋆⋆ In (van der Torre and Tan, 1997a), we defined a set of monadic obligations called the actual obligation set, instead of dyadic obligations of the active obligation set. Unfortunately, that definition is flawed because with OSD = (OBL,FACTS)=($\{O(q \mid \top)\}, \{p\}$) the active obligation set contains $O_a q$ as well as $O_a(p \wedge q)$. The latter derivation is obviously undesirable.

the set of fulfilled norms Δ_f. The following theorem states that if we consider the corresponding obligation system OSD of the normative system NSD, then the preference ordering induced by OSD on PDL worlds corresponds to the preference ordering induced by NSD on DIO(DE)^2 models.[‡]

THEOREM 1. First we define a Kripke world representation of the ordering on potential diagnoses in DIO(DE)^2, and then we define a mapping from NSD of DIO(DE)^2 to OSD of PDL. Let $\leq_M$ be a relation on models of ND $\cup$ FACTS of NSD such that $M_1 \leq_M M_2$ if and only if there are potential diagnosis $\Delta_1 = (\Delta_f, \Delta_v)$ and $\Delta_2 = (\Delta'_f, \Delta'_v)$ such that $M_1 \models \text{CONTEXT}_{\Delta 1}$, $M_2 \models \text{CONTEXT}_{\Delta 2}$, and $(\Delta_f, \Delta_v) \leq (\Delta'_f, \Delta'_v)$. Moreover, let M_K be the Kripke model $\langle W, \leq_w, V\rangle$ representation of this ordering on models, i.e. with w and M identical valuations and that preserves the ordering $\leq_M$ in $\leq_w$. Finally, let τ be the mapping of a normative system to be diagnosed NSD of DIO(DE)^2 to an obligation system to be diagnosed OSD such that for each norm $n_i \in$ NORMS where the norm description ND_F contains the formula $\neg V(n_i) \leftrightarrow (\beta_i \rightarrow \alpha_i) \wedge F(n_i) \leftrightarrow (\beta_i \wedge \alpha_i)$ there is an obligation $O(\alpha_i \mid \beta_i) \in$ OSD.

Assume that we have an NSD such that $(\Delta'_f, \Delta'_v) \not\leq (\Delta_f, \Delta_v)$ if and only if there is a norm $n \in$ NORMS such that $n \in \Delta_f$ and $n \in \Delta'_v$. Under this assumption, the models $\langle W, \leq_w, V\rangle$ are models of MAO of OSD = τ(NSD), and there are no models $\langle W, \leq'_w, V\rangle$ of MAO with $\leq_w \subset \leq'_w$.

Proof. If there is a norm $n \in$ NORMS such that $n \in \Delta_f$ and $n \in \Delta'_v$ iff $(\Delta'_f, \Delta'_v) \not\leq (\Delta_f, \Delta_v)$, i.e. $M_1 \not\leq_M M_2$, then according to the mapping τ there is an obligation $O(\alpha \mid \beta) \in$ OSD such that we have $w_1 \not\leq_w w_2$, where M_1, M_2 and w_1, w_2 have identical valuations. Hence, the theorem follows directly for OBL instead of MAO from the assumption and the mapping τ. We have to proof that the models of MAO are exactly the models of OBL, i.e. $M \models_{\text{PDL}} \text{OBL}$ if and only if $M \models_{\text{PDL}} \text{AO}$. This follows from the fact that the ordering $\leq_w$ is transitive, because for any two worlds w_1, w_2 we can find sets $W_1 \cup W_2 = W$ with $w_1 \in W_1$, $w_2 \in W_2$, and for all $v_1 \in W_1$ and $v_2 \in W_2$ we have $v_1 \not\leq v_2$. □

The following example illustrates Theorem 1 in a non-trivial case. The obligation system contains an obligation whose antecedent is not a tautology (otherwise, the theorem follows directly from the assumption and the mapping τ). Moreover, the example illustrates the use of logical derivations in PDL. It is a translation of Example 2 based on the mapping in Theorem 1.

[‡] The distinction between models and worlds explains the following distinction between DIO(DE)^2 and PDL. In DIO(DE)^2 we have that if α is implied by the facts, then α is also contextually obliged. The PDL counterpart is that $O(\alpha \mid \beta) \rightarrow O(\alpha \wedge \beta \mid \beta)$ is a theorem of the logic PDL.

EXAMPLE 4 (Transitivity,continued) Consider the obligation system of Example 2 of the two norms 'party A has to deliver the goods (d)' and 'if party A delivers the goods, then it has to give notice of the expected arrival date to party B in advance (n)', together with the fact that 'party A did not give notice'.

- OBL = $\{O(d \mid \top), O(n \mid d)\}$,
- FACTS = $\{\neg n\}$.

In the logic PDL $O(d \wedge n \mid \top)$ is derivable, which is the only derivable obligation (except logical equivalences) with antecedent $\top$. The set of active obligations contains $\{O(d \mid \top), O(d \wedge n \mid \top)\}$, which is a minimal active set. The two potential diagnosis are $\{O(d \mid \top), O(d \wedge n \mid \top)\}$ and $\{O(d \wedge n \mid \top)\}$ and the two contexts of these diagnoses imply respectively $\{\neg d, \neg n\}$ and $\{d, \neg n\}$. Finally, we compare the analysis in PDL with the analysis in DIO(DE)^2 in Example 2. The diagnosis $\{O(d \mid \top), O(d \wedge n \mid \top)\}$ corresponds to $(\Delta_f, \Delta_v) = (\emptyset, \{n_1\})$ and $\{O(d \wedge n \mid \top)\}$ corresponds to $(\Delta_f, \Delta_v) = (\{n_1\}, \{n_2\})$, in the sense that their contexts imply the same factual fragment. In PDL there is a minimal and a maximal diagnosis, whereas in DIO(DE)^2 the two diagnoses are both minimal.

Theorem 1 is in general not valid when its assumption is not satisfied. The following example illustrates that the distinction is caused by the treatment of active obligations.

EXAMPLE 5. Consider the following normative system of DIO(DE)^2.

- NORMS = $\{n_1, n_2\}$,
- $\text{ND}_F = \left\{ \begin{array}{l} (\neg V(n_1) \leftrightarrow ((p \vee q) \rightarrow p)) \wedge (F(n_1) \leftrightarrow p), \\ (\neg V(n_2) \leftrightarrow ((p \vee \neg q) \rightarrow p)) \wedge (F(n_2) \leftrightarrow p) \end{array} \right\}$,
- FACTS = $\{\top\}$.

and its mapping to OSD.

- OBL = $\{O(p \mid p \vee q), O(p \mid p \vee \neg q)\}$,
- FACTS = $\{\top\}$.

The logic PDL derives the obligation $O(p \mid \top)$, which is a minimal active obligation set. Hence, the models of MAO only distinguish between p and $\neg p$ worlds. The ordering on potential diagnosis of DIO(DE)^2 prefers the diagnosis $(\{n_1, n_2\}, \emptyset)$ to the incomparable $(\emptyset, \{n_1\})$ and $(\emptyset, \{n_2\})$. Hence, the related ordering on models distinguishes between p, $\neg p \wedge q$ and $\neg p \wedge \neg q$ worlds. Moreover, consider the following normative system of DIO(DE)^2.

- NORMS = $\{n_1, n_2\}$,

– $\text{ND}_F = \left\{ \begin{array}{l} (\neg V(n_1) \leftrightarrow (q \rightarrow p)) \wedge (F(n_1) \leftrightarrow (p \wedge q)), \\ (\neg V(n_2) \leftrightarrow (\neg q \rightarrow p)) \wedge (F(n_2) \leftrightarrow (p \wedge \neg q)) \end{array} \right\}$,
– FACTS = $\{\top\}$.

and its mapping to OSD.

– OBL = $\{O(p \mid q), O(p \mid \neg q)\}$,
– FACTS = $\{\top\}$.

The logic PDL does not derive any obligation with tautological antecedent, so the minimal active obligation set is empty. Hence, the models of MAO do not distinguish between worlds. However, there are four potential diagnoses in DIO(DE)^2: $(\{n_1\}, \emptyset)$, $(\emptyset, \{n_1\})$, $(\{n_2\}, \emptyset)$ and $(\emptyset, \{n_2\})$.

In this paper, we proposed two ways to formalize normative diagnosis and decisions: the special purpose formalism DIO(DE)^2 and the deontic logic PDL with additional principles. We end with a small comparison of both approaches, from the perspective of knowledge representation. First, an advantage of the formalization in PDL is its potential to be embedded in more general accounts of normative reasoning.⋆ DIO(DE)^2 is a special purpose formalism, which can only be used for determining normative diagnoses and decisions and not for other aspects of normative problem solving, like classification, detection and resolution of normative conflicts, etc. PDL, on the other hand, is a full-fledged deontic logic in which for example permissions can be expressed, and conflicts of obligations can be detected and resolved. An application that goes beyond normative diagnosis requires a full representation of a legal text, including all its deontic features, and is better formalized in the latter. On the other hand, an advantage of DIO(DE)^2 over a full-fledged deontic logic like PDL is that DIO(DE)^2 only needs a a fragment of deontic logic, i.e. no permissions, negated obligations, disjunctions of obligations, nested obligations. This makes the formal system computationally less complex. For example, for OIOOE^2 there are simple and efficient algorithms based on so-called conflict sets, see (Reiter, 1987). Finally, we think that a formalization in DIO(DE)^2 with its V and F predicates is often conceptually more clear than deontic logic, because it is built from first principles. For example, consider the different degrees of violation in Example 3. In DIO(DE)^2, different violation predicates or different sub-norms can easily and intuitively formalize them. In deontic logic, however, they give rise to so-called contrary-to-duty obligations, see (Forrester, 1984). They can also be formalized in preference-based semantics, but their reading is counterintuitive. For

⋆ The previous examples illustrate that we cannot take any deontic logic and add additional principles to it, because the logic has to derive a reasonable set of actual obligations. For example, in so-called Standard Deontic Logic SDL there is a theorem $O\alpha \rightarrow O(\alpha \vee \beta)$, which introduces an infinite set of obligations from a single premise. With such a base logic, it is difficulty to see how diagnoses can be defined. For a further discussion on this issue, see (Ramos & Fiadeiro, 1996, 1998; van der Torre et al., 1997).

example, in the libel article of Example 3 the two sub-norms are formalized by 'you should not insult someone,' and 'if you insult someone, then you should do it in private' (Tan & van der Torre, 1994b).

3.2. RELATED RESEARCH

The relation between qualitative decision theory and deontic logic has been observed by several researchers, see e.g. (Pearl, 1993; Boutilier, 1994; Lang, 1996). Pearl investigates a decision-theoretic account of conditional ought statements. He argues that the resulting account forms a sound basis for qualitative decision theory, thus providing a framework for qualitative planning under uncertainty. Boutilier develops a logic of qualitative decision theory in which the basic concept of interest is the notion of *conditional preference*. Boutilier writes $I(\alpha \mid \beta)$, read 'ideally α given β,' to indicate that the truth of α is preferred, given β. This holds exactly when α is true at the most preferred of those worlds satisfying β. Boutilier observes that from a practical point of view, $I(\alpha \mid \beta)$ means that if the agent (only) knows β, and the truth of β is fixed (beyond his control), then the agent ought to ensure α. Otherwise, should $\neg\alpha$ occur, the agent will end up in a less than desirable β-world. Boutilier also observes that the statement can be *roughly* interpreted as 'if β, do α'. Moreover, Boutilier observes that the conditional logic of preferences he proposed is similar to the (purely semantic) deontic logic put forth by Hansson (1971). He concludes that one may simply think of $I(\alpha \mid \beta)$ as expressing a conditional obligation to see to it that α holds if β does. Thomason and Horty (1996) and Lang (1996) also observe the relation between qualitative decision theory and deontic logic when they develop the foundations for qualitative decision theory.

It has also been observed in qualitative decision theory that decision-theoretic reasoning can be considered as an extension of deontic reasoning. The simplest definition of goals is in accordance with the general maxim 'do the best thing possible consistent with your knowledge'. This maximum can be viewed as a strategy for rational agent behavior that is determined by norms. This maximum is an extra principle on top of deontic logic that explains how norms could influence behavior. Boutilier (1994) dubbed such goals CK goals, because they seem correct when an agent has *Complete Knowledge* of the world (or at least of uncontrollable atoms).

4. Conclusions

In this paper we discussed DIO(DE)^2, the DIagnostic and DEcision-theoretic framework for DEontic reasoning. We used the framework to illustrate the distinction between diagnosis theory and (qualitative) decision theory. A crucial distinction between the two theories is their perspective on time. Diagnosis theory reasons about incomplete knowledge and only considers the past. It distinguishes between violations and non-violations. Qualitative decision theory reasons about decision variables and considers the future. It distinguishes between fulfilled obligations and

unfulfilled obligations. Moreover, in this paper we used our preference-based deontic logic PDL to show how deontic logic can be used as a component in normative diagnosis theory as well as qualitative decision theory.

Acknowledgements

This work was partially supported by the Esprit WG 8319 (MODELAGE). Thanks to an anonymous referee for useful comments on an earlier version of this article.

References

Anderson, A. 1958. A reduction of deontic logic to alethic modal logic. *Mind* **67**, 100–103.

Bench-Capon, T. 1994. Deontic logic: Who needs it? In *Proceedings of the Workshop 'Artificial Normative Reasoning' of the Eleventh European Conference on Artificial Intelligence (ECAI'94)*, pp. 69–78.

Boutilier, C. 1994. Toward a logic for qualitative decision theory. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, San Francisco, (CA): Morgan Kaufmann, pp. 75–86.

Dean, T. & Wellman, M. 1991. *Planning and Control*, San Mateo: Morgan Kaufmann.

Doyle, J. & Wellman, M. 1991. Preferential semantics for goals. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-91)*, pp. 698–703.

Doyle, J. 1980. A model for deliberation, action and introspection. Technical Report AI-TR-581, MIT AI Laboratory.

Forrester, J. 1984. Gentle murder, or the adverbial samaritan. *Journal of Philosophy* **81**, 193–197.

Gordon, T. 1995. *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*, Kluwer.

Hage, J. 1996. A theory of legal reasoning, and a logic to match, *Artificial Intelligence and Law* **4**, 199–273.

Hansson, B. 1971. An analysis of some deontic logics. In Hilpinen, R. (ed.), *Deontic Logic: Introductionary and Systematic Readings*. D. Reidel Publishing Company: Dordrecht, Holland, pp. 121–147.

Jones, A. & Sergot, M. 1992. Deontic logic in the representation of law: Towards a methodology. *Artificial Intelligence and Law* **1**, 45–64.

Jones, A. & Sergot, M. 1993. On the characterisation of law and computer systems: The normative systems perspective. In Meyer, J. & Wieringa, R. (eds.), *Deontic Logic in Computer Science*, John Wiley and Sons.

Lang, J. 1996. Conditional desires and utilities – An alternative approach to qualitative decision theory. In *Proceedings of the Tenth European Conference on Artificial Intelligence (ECAI'96)*, pp. 318–322.

McCarty, L. 1994. Modalities over actions: 1. Model theory. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, San Francisco (CA): Morgan Kaufmann, pp. 437–448.

Pearl, J. 1993. From conditional oughts to qualitative decision theory. In Heckerman, D. & Mamdani, A. (eds.), *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence (UAI-93)* San Mateo (CA): Morgan Kaufmann, pp. 12–20.

Prakken, H. & Sartor, G. 1996. A system for defeasible argumentation, with defeasible priorities. *Artificial Intelligence and Law* **4**, 331–368.

Ramos, P. & Fiadeiro, J. 1996. Diagnosis in organisational process design. Technical report, Department of Informatics, Faculty of Sciences, University of Lisbon.

Ramos, P. & Fiadeiro, J. 1998. A deontic logic for diagnosis of organisational process design. In *Proceedings of the Fourth International Workshop on Deontic Logic in Computer Science (Δeon-98)*, To appear.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial Intelligence* **32**, 57–95.

Smith, T. 1994. *Legal Expert Systems: Discussion of Theoretical Assumptions*. Ph.D. Dissertation, University of Utrecht.

Tan, Y.-H. & van der Torre, L. 1994a. DIODE: deontic logic based on diagnosis from first principles. In *Proceedings of the Workshop 'Artificial Normative Reasoning' of the Eleventh European Conference on Artificial Intelligence (ECAI'94)*, pp. 21–39.

Tan, Y.-H. & van der Torre, L. 1994b. Multi preference semantics for a defeasible deontic logic. In *Legal Knowledge-Based Systems. The Relation with Legal Theory. Proceedings of the JURIX'94*, pp. 115–126. Lelystad: Koninklijke Vermande.

Tan, Y.-H. & van der Torre, L. 1994c. Representing deontic reasoning in a diagnostic framework. In *Proceedings of the Workshop on Legal Applications of Logic Programming of the Eleventh International Conference on Logic Programming (ICLP'94)*, pp. 138–150.

Tan, Y.-H. & van der Torre, L. 1996. How to combine ordering and minimizing in a deontic logic based on preferences. In *Deontic Logic, Agency and Normative Systems. Proceedings of the Δeon'96. Workshops in Computing*, Springer Verlag, pp. 216–232.

Thomason, R. & Horty, R. 1996. Nondeterministic action and dominance: Foundations for planning and qualitative decision. In *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK'96)*, Morgan Kaufmann, pp. 229–250.

van der Torre, L. & Tan, Y.-H. 1997a. Distinguishing different roles in normative reasoning. In *Proceedings of the Sixth International Conference on AI and Law (ICAIL'97)*, ACM Press, pp. 225–232.

van der Torre, L. & Tan, Y.-H. 1998. Prohairetic Deontic Logic (PDL). In *Logics in Artificial Intelligence*, LNAI 1489, pp. 77–91, Springer.

van der Torre, L., Ramos, P., Fiadeiro, J.-L., & Tan, Y. 1997. The role of diagnosis and decision theory in normative reasoning. In *Proceedings of the Third Workshop on Formal Models of Agents (Modelage'97)*.

von Wright, G. 1983. *Practical Reason*. Blackwell: Oxford.