Distinguishing Different Roles in Normative Reasoning

Leendert W.N. van der Torre*
Max Planck Institute for Computer Science
Im Stadwald
D-66123 Saarbrücken
torre@mpi-sb.mpg.de

Yao-Hua Tan*
EURIDIS
Erasmus University Rotterdam
PO Box 1738, 3000 DR Rotterdam
ytan@euridis.fbk.eur.nl

Abstract

In this paper we introduce the DIagnostic and DEcision-theoretic framework for DEontic reasoning DIO(DE)^2. The framework DIO(DE)^2 formalizes reasoning about violations and goals. We use this framework to discuss two theories of normative reasoning, 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, we discuss the relation between the two theories of normative reasoning and deontic logic. The theories formalize reasoning *with* norms, and they are thus different from deontic logic, that formalizes reasoning *about* norms.

1 Introduction

There is a discussion in AI and law literature whether deontic logic should be used to formalize legal reasoning (and normative reasoning in general). Jones and Sergot [JS92, JS93] argue extensively and convincingly 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 [McC94] 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 [BC94] 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.

In this paper we argue that normative reasoning is more than deontic logic. Deontic logic tells you 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. 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. In this paper we discuss two formalisms that can be used to analyze two different types of aspects of how norms affect behavior, namely diagnosis theory and qualitative decision theory, represented in Figure 1. Diagnosis theory reasons about violations, and check systems against given principles. In particular, it reasons about the past with incomplete knowledge (if everything is known than a diagnosis is completely known). Diagnosis theory formalizes the reasoning of a judge 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 qualitative decision theory is that it is goal oriented reasoning, usually for planning problems. Moreover, it combines reasoning about goals with uncertainty. This reasoning is based on the application of strategies, which can be considered as qualitative versions of the 'maximum utility' criterion.

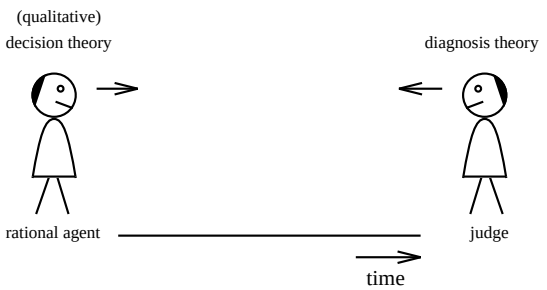

Figure 1: Reasoning with norms

In this paper we also argue that deontic logic can be used

*This work was partially supported by the Esprit WG 8319 (MODELAGE)

as a component in diagnosis theory as well as qualitative decision theory, because logical relations between obligations are an essential component of any formalism that explains the effect of norms on behavior. Actually, we even argue for the stronger claim that diagnosis theory as well as qualitative decision theory can be viewed as extensions of deontic logic. 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. There are several structural similarities between deontic logic and the logics developed for diagnosis and qualitative decision theory, see e.g. [Bou94, Lan96]. However, the distinction between deontic logic and the different perspectives raises several important questions.

Norms and dedicated theories. 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. 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?

Norms and preferences [Lan96]. 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). 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. 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 decider differ from an ordinary decider, if any?

Norms and obligations. Diagnosis theory and qualitative decision theory reason about actual behavior, whereas deontic logic reasons about ideal behavior. In deontic logic we only reason about facts when we consider which absolute obligations can be derived from a set of conditional obligations and facts by so-called factual detachment. Should we distinguish norms from the obligations derivable from the norms and a set of facts? What is the role of analogues of factual detachment in diagnosis theory and qualitative decision theory?

In this paper we introduce the DIagnostic and DEcision-theoretic framework for DEontic reasoning DIO(DE)^2. We use this framework to discuss the distinction between diagnostic reasoning and decision-theoretic reasoning. DIO(DE)^2 contains two main ingredients. First, it contains representations of violations to formalize the reasoning about violations of the DIagnostic framework for DEontic reasoning DIODE [TvdT94a, TvdT94b], Reiter's theory of diagnosis [Rei87] applied to normative systems. Second, it contains representations of fulfilled obligations, which make it possible to formalize reasoning about goals. Moreover, we use the DIO(DE)^2 framework to discuss the relation between the two theories of normative reasoning and deontic logic.

The layout of this paper is as follows. In Section 2 we discuss Reiter's theory of diagnosis, the adaptation of that theory to deontic systems DIODE by using obligations to represent the ideal behavior of a system, and several adaptations of DIODE. In Section 3 we discuss qualitative decision theory and DIO(DE)^2. In Section 4 we discuss the relation between DIO(DE)^2 and deontic logic.

2 Diagnosis theory

In this section we 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.

2.1 Reiter's theory of diagnosis

The contribution of Reiter to diagnosis theory is widely accepted. His *consistency-based approach* [Rei87] 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* $\Delta \subseteq$ 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* $\text{CONTEXT}_\Delta \models \Pi$.

2.2 Diagnostic framework for deontic reasoning DIODE

The DIagnostic framework for DEontic reasoning DIODE introduced in [TvdT94a, TvdT94b] 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 other hand, since diagnosis reasons about violations and deontic logic is useful to model situations where violations are important [JS92], 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 ND. 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.

Definition 2 (DIODE) *A* normative system *is written as a tuple* NS = (NORMS, ND) *with:*

1. NORMS, *a finite set of constants* $\{n_1, \ldots, n_k\}$ *denoting* norms,
2. ND, *the* norms description, *a set of first-order* $\mathcal{L}_V$ *sentences* $\neg V(n_i) \rightarrow (\beta \rightarrow \alpha)$ *denoting* obligations.

A normative system to be diagnosed *is written as a tuple* NSD = (NORMS, ND, FACTS) *with:*

1. NS = (NORMS, ND), *a normative system, and*
2. FACTS, *a set of first-order* $\mathcal{L}$ *sentences that describe the facts.*

Let NSD = (NORMS, ND, FACTS) *be a normative system to be diagnosed. A* minimal violated-norm set Δ *of* NSD *is a minimal (with respect to set inclusion) 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. The set of contextual obligations *of a minimal violated-norm set* Δ *of a normative system to be diagnosed* NSD *is* $\text{CO}_\Delta = \{\alpha \mid \alpha \in \mathcal{L}, \text{CONTEXT}_\Delta \models \alpha\}$.

Obligations are represented in DIODE analogously to the way they are represented in Anderson's reduction of so-called Standard Deontic Logic (SDL) to alethic modal logic. SDL is a normal modal system of type KD according to the Chellas classification [Che80], in which the modal sentence Op is read as 'p ought to be (done).' It satisfies, besides the propositional tautologies, **K**: $O(\alpha \rightarrow \beta) \rightarrow (O\alpha \rightarrow O\beta)$, which states that modus ponens holds within the scope of the modal operator, and **D**: $\neg(O\alpha \wedge O\neg\alpha)$, which states that dilemmas are inconsistent. Anderson [And58] showed that SDL can be expressed in alethic modal logic by the translation $O\alpha =_{\text{def}} \Box(\neg V \rightarrow \alpha)$, in which V is the so-called violation constant (not a propositional variable!), together with the axiom **D**: $\Diamond\neg V$ (as usual, $\Diamond\alpha =_{\text{def}} \neg\Box\neg\alpha$). In SDL, a conditional obligation can be represented by $\beta \rightarrow O\alpha$ or by $O(\beta \rightarrow \alpha)$. The latter is according to the Anderson schema similar to $O(\beta \rightarrow \alpha) =_{\text{def}} \Box(\neg V \rightarrow (\beta \rightarrow \alpha))$. In spite of the analogy in the way obligations are represented, there are also two important distinctions between the representation of obligations in DIODE and Anderson's reduction. First, in Anderson's reduction every deontic formula is preceded by a box $\Box$. Semantically, in diagnosis theory distinct *models* represent distinct situations, whereas in a modal system distinct *worlds* within a model represent distinct situations. Second, in Anderson's reduction there is only one violation constant. For a further discussion see [TvdT94a]. In spite of the analogy in the representation of obligations in DIODE and Anderson's reduction, DIODE is quite different from a deontic logic. On the one hand DIODE is more than a deontic logic, because the parsimony principle adds the assumption that the set of violations of obligations is minimal. This assumption is based on the idea that people tend to comply with norms, which is an empirical assumption about the behavior of people, and which has clearly nothing to do with the logic of norms itself. On the other hand one could argue that DIODE is less than a deontic logic, because, if ND would be a deductively closed set of sentences, then the DIODE counterpart of the formula $p \rightarrow Op$ would be contained in every ND. Clearly, $p \rightarrow Op$ is not an intuitive deontic theorem, and the counterpart of this formula is also not valid in Anderson's reduction due to his box operator. Although these counter-intuitive theorems do not occur in DIODE, because ND is not deductively closed, we give another formulation at the end of this paper of DIODE in the logic 2DL, which gives a better representation of the deontic logic component in DIODE.

2.3 Adaptation 1: distinguishing potential and minimal diagnoses (Ramos and Fiadeiro)

Ramos and Fiadeiro [RF96b] observe that DIODE 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, sometimes the assumption 'guilty until proven innocent' is needed. They therefore distinguish between potential diagnoses (any subset of NORMS such that CONTEXT_Δ is consistent) and minimal and maximal violated-norm sets. Moreover, they observe a lack of so-called fault knowledge in DIODE. Fault knowledge (see e.g. [dKMR90]) describes the consequences of broken components. For example, $\beta \wedge Ab(c) \rightarrow \alpha$ describes the consequences α of faulty component c in circumstances β. Hence, with fault knowledge from the abnormality of a component new information can be derived. If the rules from the system description SD are represented by $\beta \wedge \neg Ab(c) \rightarrow \alpha$, then there is no fault knowledge. The maximal violated-norm set is in that case simply the set of all components. Obviously, for any reasonable definition of a maximal violated-norm set, fault knowledge has to be added.[1] When a set of norms is formalized in Ramos and Fiadeiro's variant of DIODE, the following two assumptions (see [RF96b]) are made to incorporate fault knowledge.

Assumption 1 As a rule, each (conditional) obligation of a premise set corresponds to a separate norm n_i. If a set of obligations is formalized in DIODE, then this set of obligations is translated to a set of norms.

Assumption 2 Every norm description describes an obligation completely. Thus, a conditional obligation 'α

[1] As remarked in [dKMR90], with the representation of fault knowledge it is no longer possible to compute all consistent sets of normal and abnormal components based on minimal diagnosis, because not all supersets of minimal sets are consistent.

ought to be (done), if β is (done)' is represented in DIODE by the norm description $\neg V(n_i) \leftrightarrow (\beta \rightarrow \alpha)$. The conditional obligation can be read in DIODE as 'if the norm n_i is not violated, then and only then if β is (done) then α is (done).' The sentence is logically equivalent with $V(n_i) \leftrightarrow (\beta \wedge \neg\alpha)$, which explains why we call $V(n_i)$ a *violation constant* (although, strictly speaking, it is not a constant).

With these two assumptions we can distinguish minimal and maximal violated-norm sets (called benevolent and exigent diagnosis by Ramos and Fiadeiro [RF96a, RF96b]). Two aspects of the diagnostic approach are explicitly distinguished. 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 difference 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.

2.4 Adaptation 2: applicable norms

In DIODE, there is no distinction between fulfilling a conditional obligation, and inapplicability of a conditional obligation. For example, for an obligation 'α ought to be (done) if β is (done)' we have $\neg V(n) \leftrightarrow (\beta \rightarrow \alpha)$, which is logically equivalent with $\neg V(n) \leftrightarrow ((\beta \wedge \alpha) \vee \neg\beta)$. The following definition shows how we can add applicability information. For example, for an obligation 'α ought to be (done) if β is (done)' we have $\neg V(n) \leftrightarrow (\beta \rightarrow \alpha) \wedge A(n) \leftrightarrow \beta$. Thus, the underlying logic is extended with an applicability predicate similar to the violation predicate. We call the system DIODE-A. The use of DIODE-A consists of two steps. First we determine the applicable obligations by minimizing the $A(n)$. Second, for applicable obligations we can have minimal or maximal violated-norm sets.

Definition 3 (DIODE-A) *A* normative system *is written as a tuple* $\text{NS} = (\text{NORMS}, \text{ND}_A)$ *where* ND_A*, the* norms description*, is a set of* conditional obligations

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

Let $\text{NSD} = (\text{NORMS}, \text{ND}_A, \text{FACTS})$ *be a normative system to be diagnosed. The* active norms Δ_a *of* NSD *is a minimal subset of* NORMS *such that*

$$\text{ND}_A \cup \text{FACTS} \cup \{A(n_i) \mid n_i \in \Delta_a\} \cup \{\neg A(n_i) \mid n_i \in \text{NORMS} - \Delta_a\}$$

is consistent. A potential diagnosis Δ *of* NSD *is a subset of some* Δ_a *of* NSD *such that*

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

is consistent. A minimal (maximal) violated-norm set Δ *of* NSD *is a minimal (maximal) subset of some* Δ_a *of* NSD *such that* CONTEXT_Δ *is consistent.*

The following example illustrates DIODE-A.

Example 1 *Consider the normative system of the obligation 'if an order form is send (o), then a copy ought to be stored (c).'*

- $\text{NORMS} = \{n_1\}$,
- $\text{ND}_A = \{(\neg V(n_1) \leftrightarrow (o \rightarrow c)) \wedge (A(n_1) \leftrightarrow o)\}$,
- $\text{FACTS} = \emptyset$.

The set of active norms Δ_a *is empty, thus there is no potential diagnosis which contains the norm* n_1*. In particular, the only maximal violated-norm set is the empty set. Moreover, consider the following normative system of the two obligations '*p_1 *ought to be done if* q *is done' and '*p_2 *ought to be done if* $\neg q$ *is done.'*

- $\text{NORMS} = \{n_1, n_2\}$,
- $\text{ND}_A = \left\{ \begin{array}{c} (\neg V(n_1) \leftrightarrow (q \rightarrow p_1)) \wedge (A(n_1) \leftrightarrow q), \\ (\neg V(n_2) \leftrightarrow (\neg q \rightarrow p_2)) \wedge (A(n_2) \leftrightarrow \neg q) \end{array} \right\}$,
- $\text{FACTS} = \emptyset$.

Given the tautology $q \vee \neg q$*, we have for two minimal active sets* $\Delta_a = \{n_1\}$ *and* $\Delta_a = \{n_2\}$*. Finally, consider the following normative system of the two obligations '*p *ought to be done if* q *is done' and '*q *ought to be done.'*

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

The minimal active set is $\Delta_a = \{n_2\}$*. The first norm is not applicable, because there is no transitivity of the rules.*

The following example illustrates that DIODE does not suffer from the contrary-to-duty paradoxes of deontic logic like the notorious Good Samaritan, Chisholm and Forrester paradoxes. This is no surprise, because DIODE does not tell which norms follow from a set of norms.

Example 2 (Chisholm paradox) *Consider the following normative system of the Chisholm set 'a certain man should go to the assistance of his neighbors' (a), 'if the man goes then he should tell' (t), 'if the man does not go then he should not tell'* $(\neg t)$ *and he does not go* $(\neg a)$*.*

- $\text{NORMS} = \{n_1, n_2, n_3\}$,
- $\text{ND}_A = \left\{ \begin{array}{c} (\neg V(n_1) \leftrightarrow a) \wedge (A(n_1) \leftrightarrow \top), \\ (\neg V(n_2) \leftrightarrow (a \rightarrow t)) \wedge (A(n_2) \leftrightarrow a) \\ (\neg V(n_3) \leftrightarrow (\neg a \rightarrow \neg t)) \wedge (A(n_3) \leftrightarrow \neg a) \end{array} \right\}$,
- $\text{FACTS} = \{\neg a\}$.

The minimal active set is $\Delta_a = \{n_1, n_3\}$*. The first norm is violated and part of every potential diagnosis. The second norm is not applicable and not part of any violated-norm set. The third norm is applicable. It is not part of the minimal violated-norm set, but it is part of the maximal violated-norm set.*

3 Qualitative decision theory

Pearl [Pea93] 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 [Bou94] developed a logic of qualitative decision theory in which the basic concept of interest is the notion of *conditional preference*. Boutilier writes $I(\alpha|\beta)$, read 'ideally α given β,' to indicate that the truth of α is

preferred, given β. This holds exactly when α is true at each of the most preferred of those worlds satisfying β. Boutilier observes that from a practical point of view, $I(\alpha|\beta)$ 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) proposal put forth by Hansson [Han71]. He concludes that one may simply think of $I(\alpha|\beta)$ as expressing a conditional obligation to see to it that α holds if β does. Thomason and Horty [TH96] and Lang [Lan96] also observe the relation between qualitative decision theory and deontic logic when they develop the foundations for qualitative decision theory.

Boutilier [Bou94] introduces a simple model of action and ability. The atomic propositions are partitioned into *controllable* propositions, atoms over which the agent has direct influence, and *uncontrollable* propositions. He ignores the complexities required to deal with effects, preconditions and such, in order to focus attention on the structure and interaction of ability and goal determination. The consequence of this lack of an action model is that we should think of a rule as an *evidential rule* rather than a *causal rule*. Moreover, Boutilier observes the implicit temporal aspect here; propositions should be thought of as *fluents*. We can avoid an explicit temporal representation by assuming that preference is solely a function of the truth values of fluents. Lang [Lan96] calls controllable and uncontrollable propositions respectively decision variables and parameters. Moreover, he argues that it is necessary to distinguish not only between desires (goals) and knowledge as in [Bou94] but also between background factual knowledge (which tells which worlds are physically impossible) and contingent knowledge (which tells which of the physically possible worlds can be the actual states of affairs). This last distinction was taken from [vdT94].

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 [Bou94] dubbed such goals CK goals, because they seem correct when an agent has *Complete Knowledge* of the world (or at least of uncontrollable atoms). But Boutilier also shows that CK-goals do not always determine the best course of action if an agent's knowledge is *incomplete*. In such a case we could use for example Wald's pessimistic strategy of maximizing the minimum return (see e.g. [DP95, Lan96]).

Boutilier only considers the single agent case, in which an agent reasons about his own goals and looks for values of his decision variables. In a multi agent system, an agent also reasons about the other agents' behavior. For example, if you approach a square and your light is green, and another car approaches from the left where the light is red, then you assume that the other car will stop. This additional assumption cannot be explained by a deontic logic. McCarty [McC94] 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.

3.1 Diagnostic and decision-theoretic framework DIO(DE)^2

A theory of diagnosis like DIODE 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 is the DIagnostic and DEcision-theoretic framework for DEontic reasoning that extends DIODE. It 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. DIO(DE)^2 has fulfilled-norm constants (F). In the following definition of DIO(DE)^2, for an obligation 'α should be (done) if β is (done)' we have besides $\neg V(n) \leftrightarrow (\beta \rightarrow \alpha)$ also $F(n) \leftrightarrow (\beta \wedge \alpha)$.

Definition 4 (DIO(DE)^2) *A* normative system *is written as 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*

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

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

We minimize the applicable norms by minimizing the relation $(\Delta_f, \Delta_v) \leq (\Delta'_f, \Delta'_v)$. The following example illustrates DIO(DE)^2 and compares it with DIODE-A.

Example 3 (Transitivity) *Consider the following normative system of the two obligations 'p ought to be done if q is done' and 'q ought to be done.'*

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

The two potential diagnoses are $(\Delta_f, \Delta_v) = (\emptyset, \{n_2\})$ *and* $(\Delta_f, \Delta_v) = (\{n_2\}, \{n_1\})$. *As illustrated in Example 1 in* DIODE-A *the minimal active set is* $\Delta_a = \{n_2\}$. *The two systems do not behave similarly, because in* DIO(DE)^2 *it is possible that the first obligation is violated.*[2]

[2] There is an interesting connection between the set of obligations of Example 3 and deontic detachment (or transitivity), which is written as $O(\alpha|\beta) \wedge O(\beta|\gamma) \rightarrow O(\alpha|\gamma)$, where $O(\alpha|\beta)$ represents a conditional obligation and is read as 'α ought to be (done) if β is (done).' With deontic detachment we can derive the obligation $O(p|\top)$ from the two premises $O(p|q)$ and $O(q|\top)$. Thus, if deontic detachment is valid, then the fact $\neg p$ is a violation. In DIODE-A, there is only one active set, that contains the second obligation. It is possible that this obligation is fulfilled, and there are therefore no violations. On the other hand, in DIO(DE)^2 every potential diagnosis contains violations.

We end this section with two technical remarks that will be useful in the following section. Also notice that we can distinguish two uses of a theory of diagnosis. First, we can consider a normative system to be diagnosed NSD and calculate the potential fulfilled-violated sets, the minimal sets etc. Secondly, we can consider a normative system NS and consider the mapping of facts to the different types of norm sets. The following two observations state that the fulfilled-violated sets are characterized by an ordering on models and by the set of contextual obligations.

Ordering Consider the set of models of a normative system NS = (NORMS, ND) with the unique ordering on the models $M_1 \leq M_2$ if and only if there is no norm n_i such that $M_1 \models V(n_i)$ and $M_2 \models F(n_i)$. Notice that this ordering is reflexive but not necessarily transitive. This ordering represents the mapping of facts to the different types of norm sets. It is easily seen that the fulfilled-violated set (Δ_f, Δ_v) is a potential diagnosis of NSD = (NORMS, ND, FACTS) if there is a model M such that

1. $M \models$ FACTS, for all $n_i \in \Delta_f$ we have $M \models F(n_i)$ and for all $n_i \in \Delta_v$ we have $M \models V(n_i)$, and
2. for all M' such that $M' \models$ FACTS and $M' \leq M$, we have if $M \models F(n_i)$ then $M' \models F(n_i)$ and we have if $M \models V(n_i)$ then $M' \models V(n_i)$.

Contextual obligations The fulfilled-violated sets and the set of contextual obligations both characterize a diagnosis in the sense that when you know the norm constants you know the set of contextual obligations and vice versa. Although Reiter's theory of diagnosis is primarily interested in the set of norm constants, it can equivalently be defined in terms of contextual obligations.

4 The relation with deontic logic

In this section we observe a distinction in deontic logic analogues to the distinction between diagnosis theory and qualitative decision theory. Moreover, we discuss the relation between DIO(DE)^2 and our two-phase deontic logic 2DL.

4.1 Context of justification versus context of deliberation

The distinction between the perspective of a rational agent (qualitative decision theory) and a judge (diagnosis theory) corresponds to Thomason's distinction between the context of deliberation and the context of justification [Tho81]. He distinguishes between two ways in which the truth values of deontic sentences are time-dependent. First, these values are time-dependent in the same, familiar way that the truth values of all tensed sentences are time-dependent. Second, their truth values are dependent of a set of choices or future options that varies as a function of time. If you think of deontic operators as analogous to quantifiers ranging over options, this dependency on context is a familiar phenomenon. The distinction between the context of deliberation and the context of justification follows from the second way. The context of deliberation is the set of choices when you are looking for practical advice, whereas the context of justification is the set of choices for someone who is judging you.[3]

[3] Thomason defines the context of justification in terms of the context of deliberation. At a certain point in time α is obligatory in the context of justification if and only if at some earlier point in time α was obligatory in the context of deliberation (in both cases α has the same time index). This is in our opinion too simple. We should make a distinction analogous to the distinction between revision and update to formalize the distinction, as is discussed below.

The following example discussed in [Han71] illustrates that it is important to discriminate between these two contexts, because a sentence can sometimes be interpreted differently in each of them.

Example 4 (Smoking) *Consider the obligation 'Ron ought not to smoke if he smokes.' In the context of justification the obligation is interpreted as the identification of the fact that Ron is violating a rule, whereas in the context of deliberation it is interpreted as the obligation to stop smoking. When the context is not known, it is also not known which of these two interpretations (or probably both) is meant. The two perspectives are represented in Figure 2. At the present moment in time, Ron is smoking (s). The context of justification considers the moment before the truth value of s was settled, and considers whether at that moment in the past, $\neg s$ was preferred over s. The context of deliberation considers the moment the truth value of s can be changed, and considers whether at that moment in the future, $\neg s$ will be preferred over s.*

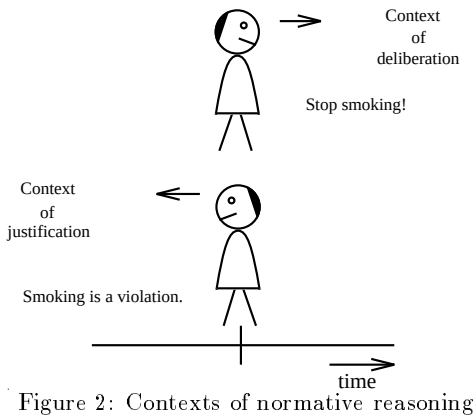

Figure 2: Contexts of normative reasoning

The consequence of the fact that there are two interpretations of the same deontic formula is that there are two distinct types of obligations. Let us call the obligations O_j and O_d, respectively. The sentence 'Ron smokes and he ought not to smoke' $s \wedge O_j \neg s$ means the identification of the fact that he is violating a rule, whereas $s \wedge O_d \neg s$ means that he should stop smoking. The two types of obligations are independent. This means that we can have $s \wedge O_j \neg s \wedge \neg O_d \neg s$ as well as $s \wedge \neg O_j \neg s \wedge O_d \neg s$. The distinction between the two interpretations of the obligation 'Ron ought to not smoke if he smokes' is as important as the distinction between Alchourrón-Gärdenfors-Makinson belief revision (or theory revision) [AGM85] and Katsuno-Mendelzon belief update [KM92] in the area of logics of belief. There is a strong analogy, because belief revision is reasoning about a non-changing world and update is reasoning about a changing world. It follows directly from Figure 2 that a similar distinction is made between respectively the context of justification and the context of deliberation, because the past is fixed, whereas the future is wide open.

4.2 The two-phase deontic logic 2DL

The two-phase preference-based deontic logic 2DL [TvdT96] is used to compare the deontic component in DIO(DE)^2 and

classical deontic logics. In the modal preference semantics of 2DL, the accessibility relation is interpreted as a betterness 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 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 [TvdT96, vdTT97b].

Definition 5 (2DL) *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 $M \models O(\alpha|\beta)$ 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'.*

We have $M \models O_\exists(\alpha|\beta)$ if and only if

1. *there is a w with $M, w \models \alpha \wedge \beta$ such that for all w' with $M, w' \models \neg\alpha \wedge \beta$, we have $w' \not\leq w$, and*
2. *there is such a world w'.*

The following definition illustrates that the modal logic 2DL can be used as the basis of a diagnosis or decision theory.[4]

Definition 6 (Deontics-based 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 actual obligation set AO *is the set of obligations (without logical equivalents):*

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

A potential diagnosis Δ is a subset of the actual obligation set AO *such that*

$$\text{CONTEXT}_\Delta = \text{OBL} \cup \text{FACTS} \cup \text{STRUCT} \cup \{\neg\alpha \mid O_a\alpha \in \Delta\} \cup \{\alpha \mid O_a\alpha \in \text{AO} - \Delta\}$$

is consistent.

The minimizing obligations $O_\exists(\alpha|\beta)$ of Definition 5 can be used in the definition of deontics-based diagnosis if we are only interested in the minimal violated-norm sets. The following theorem shows that DIO(DE)^2 corresponds to diagnosis based on the deontic logic 2DL.

[4] The actual obligation set can also be defined in the language of the deontic logic 2DL. We can use the factual detachment derivation $\beta \wedge O(\alpha|\beta) \rightarrow O\alpha$. We did not do this, because this does not buy us anything: the monadic obligations do not have any interesting properties. For example, in contrast to the dyadic obligations they are not closed under the conjunction rule, see [TvdT96].

Theorem 1 (DIO(DE)^2 **and** 2DL) *Consider the mapping of* OSD *to* DIO(DE)^2 *such that there is a norm $n_i \in$* NORMS *for each obligation $O(\alpha_i|\beta_i) \in$* OSD *and* ND_F *contains the formula $\neg V(n_i) \leftrightarrow (\beta_i \rightarrow \alpha_i) \wedge F(n_i) \leftrightarrow (\beta_i \wedge \alpha_i)$. The potential diagnosis of* OSD *are mapped on the potential diagnosis of* DIO(DE)^2.

Proof *The correspondence follows directly from the preference-based semantics. An obligation $O(\alpha\,|\,\beta)$ in* DIO(DE)^2 *is a preference of $\alpha \wedge \beta$ (fulfilled norm) over $\neg\alpha \wedge \beta$ (violated norm). This preference is defined in two steps: in the base language the fulfilled and violated norm constants are defined, and in the definition of potential diagnosis the set of applicable norms is minimized. In* 2DL, *the preference is not represented by fulfilled and violated norm constants, but defined directly in the preference-based semantics.*

A corollary of Theorem 1 is that DIO(DE)^2 is the deontic logic 2DL in which certain aspects (fulfillments and violations) are made explicit with the use of a naming convention, i.e. to use names n_i to denote norms.

5 Conclusions

In this paper we introduced 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, we used the framework to discuss the relation between the two theories and deontic logic.

There are several issues for further research. Some interesting questions have been raised in the introduction of this paper.

References

[AGM85] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, pages 510–530, 1985.

[And58] A.R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind*, 67:100–103, 1958.

[BC94] T. Bench-Capon. Deontic logic: Who needs it? In *Proceedings of workshop 'Artificial Normative Reasoning' of the Eleventh European Conference on Artificial Intelligence (ECAI'94)*, pages 69–78, Amsterdam, 1994.

[Bou94] C. Boutilier. Toward a logic for qualitative decision theory. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, pages 75–86, San Francisco, CA, 1994. Morgan Kaufmann.

[Che80] B.F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.

[dKMR90] J. de Kleer, A.K. Mackwort, and R. Reiter. Characterizing diagnosis. In *Proceedings of the National Conference on Artificial Intelligence (AAAI'90)*, pages 324–330, Boston, MA, 1990.

[DP95] D. Dubois and H. Prade. Qualitative decision theory. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*, pages 1924–1930. Morgan Kaufman, 1995.

[Han71] B. Hansson. An analysis of some deontic logics. In R. Hilpinen, editor, *Deontic Logic: Introductionary and Systematic Readings*, pages 121–147. D. Reidel Publishing Company, Dordrecht, Holland, 1971.

[JS92] A.J.I. Jones and M. Sergot. Deontic logic in the representation of law: Towards a methodology. *Artificial Intelligence and Law*, 1:45–64, 1992.

[JS93] A.J.I. Jones and M. Sergot. On the characterisation of law and computer systems: The normative systems perspective. In J.J. Meyer and R. Wieringa, editors, *Deontic Logic in Computer Science*. John Wiley & Sons, 1993.

[KM92] H. Katsuno and A.O. Mendelzon. On the difference between updating a belief base and revising it. In P. Gärdenfors, editor, *Belief Revision*, pages 183–203. Cambridge University Press, 1992.

[Lan96] J. Lang. Conditional desires and utilities - an alternative approach to qualitative decision theory. In *Proceedings of the Tenth European Conference on Artificial Intelligence (ECAI'96)*, pages 318–322, 1996.

[McC94] L.T. McCarty. Modalities over actions: 1. model theory. In *Proceedings of the Fourth International Conference on Principles of Knowledge Representation and Reasoning (KR'94)*, pages 437–448, San Francisco, CA, 1994. Morgan Kaufmann.

[Pea93] J. Pearl. From conditional oughts to qualitative decision theory. In D. Heckerman and A. Mamdani, editors, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence (UAI-93)*, pages 12–20, San Mateo, CA, 1993. Morgan Kaufmann.

[Rei87] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[RF96a] P. Ramos and J.L. Fiadeiro. A deontic logic for diagnosis of organisational process design. Technical report, Department of Informatics, Faculty of Sciences, University of Lisbon, 1996.

[RF96b] P. Ramos and J.L. Fiadeiro. Diagnosis in organisational process design. Technical report, Department of Informatics, Faculty of Sciences, University of Lisbon, 1996.

[TH96] R. Thomason and R. Horty. 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)*, pages 229–250. Morgan Kaufmann, 1996.

[Tho81] R. Thomason. Deontic logic as founded on tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 165–176. D. Reidel, 1981.

[TvdT94a] Y.-H. Tan and L.W.N. van der Torre. 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)*, pages 21–39, Amsterdam, 1994.

[TvdT94b] Y.-H. Tan and L.W.N. van der Torre. 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)*, pages 138–150, Genoa, Italy, 1994.

[TvdT96] Y.-H. Tan and L.W.N. van der Torre. 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*, pages 216–232. Springer Verlag, 1996.

[vdT94] L.W.N. van der Torre. Violated obligations in a defeasible deontic logic. In *Proceedings of the Eleventh European Conference on Artificial Intelligence (ECAI'94)*, pages 371–375. John Wiley & Sons, 1994.

[vdTT95] L.W.N. van der Torre and Y.H. Tan. Cancelling and overshadowing: two types of defeasibility in defeasible deontic logic. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI'95)*. Morgan Kaufman, 1995.

[vdTT97a] L.W.N. van der Torre and Y.H. Tan. The many faces of defeasibility in defeasible deontic logic. In D. Nute, editor, *Defeasible Deontic Logic*. Kluwer, 1997.

[vdTT97b] L.W.N. van der Torre and Y.H. Tan. Prohairetic deontic logic and qualitative decision theory. In *Proceedings of AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*, 1997. To appear.