Causal Deontic Logic

Leendert van der Torre
Department of Artificial Intelligence, Vrije Universiteit Amsterdam
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands
TORRE@CS.VU.NL

Abstract

In this paper we introduce a descriptive temporal deontic logic based on causal theories. The underlying nonmonotonic temporal logic has two distinctive properties. First, it distinguishes between observations and interventions, which among others is used to distinguish between the existence and creation of deontic states such as obligations, permissions and prohibitions. Second, its explicit causal theories lead to a descriptive or modeling perspective, that not only enables a simple and intuitive formalization of the benchmark examples of nonmonotonic temporal reasoning, but that also makes the logic a good candidate for applications in computer science.

1 Introduction

This paper is a preliminary investigation in the combination of deontic and causal reasoning. We expect that this combination is useful for the following two reasons:

- Causal reasoning was succesfully combined with decision theory in causal decision theory, and there are structural similarities between deontic logic and decision logic;
- The distinction between dynamic interventions and static observations of causal reasoning has been used to distinguish between explained and

unexplained abnormalities in nonmonotonic temporal reasoning. Analogously it may be used to distinguish between declarations and assertions in normative reasoning, between creating an obligation or permission for another agent and evaluating whether such deontic states hold [vdTT99], and finally between the power and permission to create obligations and permissions [Mak86, FS99].

Dynamic reasoning about norms has been formalized in dynamic deontic logic [Dig99, Lia98], deontic update semantics [vdTT99] and deontic deep structure models [TT98], with applications in computer security (e.g. changing access rights) and electronic commerce (e.g. the transfer of obligations and rights in international trade procedures and contracts) [KH99, TT98]. In contrast to other dynamic approaches to normative reasoning, causal deontic logic is founded on the distinction between interventions and observations. The logic introduced in this paper extends Grünwald's nonmonotonic temporal logic [Grü98] based on Pearl's causal theories [Pea96] with institutions, agents, deontic facts and institutional facts. Grünwald extensively discusses that successful nonmonotonic temporal logics take a descriptive or 'modeling' perspective, for example Sandewall's trajectory semantics [San94], Thielscher's work on causality [Thi97], and the work on action and change in the probabilistic community [DP94, Gef97]. The descriptive element of the logic – called causal theories – has to be introduced to solve the so-called frame and ramifications problems (based on persistence of fluents and indirect side-effects of actions) and their benchmark examples such as the Yale shooting problem [HM86] and the stolen car problem [Kau86]. In this paper we study the descriptive *deontic* perspective based on causal reasoning for phenomenas such as authorization, dead-lines and contrary-to-duty reasoning.

The distinction between interventions and observations can be represented by proof trees. Consider for example the institutional rule that if an agent α_1 owns file x, then he has the power to give the agent α_2 the power to grant permission to agent α_3 to read file x. The derivation below represents that being empowered is a precondition for a declaration to be effective: if agent α_1 owns the file and declares that agent α_2 has the power to grant the permission, and agent α_2 declares that agent α_3 has the permission to read the file, then the agent α_3 has the permission to read it. For simplicity we assume that all facts and actions refer to the same time moment t, and we leave the institution implicit. The distinction between being empowered (Pow) and permitted (P) is discussed later in this paper.

$$\cfrac{\cfrac{\cfrac{Own^t_{\alpha_1} x^t}{Pow^t_{\alpha_1} Pow^t_{\alpha_2} P^t_{\alpha_3} x^t}\ \text{institutional rule}}{Pow^t_{\alpha_2} P^t_{\alpha_3}, x^t}\ Declare^t_{\alpha_1} Pow^t_{\alpha_2} P^t_{\alpha_3} x^t}{P^t_{\alpha_3} x^t}\ Declare^t_{\alpha_2} P^t_{\alpha_3} x^t$$

The left hand side of this derivation contains the observables or institutional facts (that can be true or false), such as for example ownership of a file, and being empowered to change permissions about the file. The right hand side contains interventions or actions, such as declarations.

Another example of dynamic deontic reasoning is the transfer of deontic states called delegation, a type of norm ascription. A boss can delegate his obligation to his employee, and autonomous agents may exchange their obligations ("I will proofread your paper if you buy me a beer"). Task delegation is in particular important in circumstances where an agent does not know *how* to solve a subtask, because in such cases he also delegates a part of his planning problem. Thus, delegation and autonomy become a crucial part of problem solving. For example, consider a seller that delegates his obligation to ship goods to a shipper, because he does not know *how* to ship the goods. The seller cannot make a complete plan in which he fulfills his obligation to deliver goods to a buyer, he just makes a plan in which he delivers his goods on the assumption that the shipper fulfills his obligation towards him. The shipper has to decide himself how to ship the goods, and therefor has to be autonomous.

After delegating an obligation the agent may still be responsible. For example, when a seller delegates his obligation to ship goods to a shipper and the shipper violates his obligation, then the buyer will take the seller to court. It is only afterwards the seller takes the shipper to court. In other cases of delegation, such as the following one, the agent not only transfers his obligation but also his responsibility; after the transfer the agent α_1 no longer has an obligation (O).

$$\frac{O^t_{\alpha_1} p^{t+2} \quad Pow^t_{\alpha_1} O^{t+1}_{\alpha_2} p^{t+2}}{O^{t+1}_{\alpha_2} p^{t+2}, \neg O^{t+1}_{\alpha_1} p^{t+2}}\ Declare^t_{\alpha_1} O^{t+1}_{\alpha_2} p^{t+2}$$

Causal deontic logic is not only well fit for applications, it also clarifies old philosophical problems. For example, consider the well-know problem of deontic logic how to proceed once a norm is violated, called the contrary-to-duty problem. In static frameworks, a contrary-to-duty obligation is an obligation that is only in force in a sub-ideal situation. In the causal framework we can formalize that a violation *causes* a new obligation.

$$\frac{O^t_\alpha p^t \quad \neg p^t}{O^t_\alpha q^{t+1}}\ Ctd(O^t_\alpha q^{t+1} \mid O^t_\alpha p^t)$$

This paper is organized as follows. In Section 2 we introduce institutions, agents, deontic facts and institutional facts in Grünwald's nonmonotonic temporal logic and in Section 3 we discuss some examples of normative reasoning. The distinction between observations and interventions (i.e. left and right hand side of the proof trees) is represented by respectively a Ho and Do predicate. In Section 4 we sketch how this ought-to-be logic can be extended with an ought-to-do logic. In the Appendix A we repeat Grünwald's definitions for abnormality minimization.

2 Causal theories

Causal approaches distinguish between explained and unexplained abnormalities in nonmonotonic temporal reasoning. For example, in the well-known Yale shooting problem [HM86] the non-persistence of a gun being loaded is explained by firing the gun, whereas in the also widely discussed stolen car problem [Kau86] the non-persistence of a car being parked remains unexplained in the sense that it is not inferred when the car was stolen. In this section we extend a temporal logic based on causal theories with agents and deontic states such as obligations, permissions and prohibitions. We closely follow the logic proposed in Grünwald's PhD thesis [Grü98], that also gives an excellent survey on the state of the art in the area of nonmonotonic temporal reasoning.

The major choice of this extension is how we deal with persistence, because persistence of deontic states is different from persistence of brute facts. First, deontic operators are indexed by *two* temporal operators: one for the deontic operator and one for the fact the deontic operator refers to. For example, today we can have the obligation to go to a party next week, for example because we promised to do so, but next week we may have the more important obligation to save a child from drowning. In this case the obligation to go to the party is overridden or defeated by the obligation to save the child. Second, it is clear that in absence of interventions and abnormalities we desire to derive 'the light is on at moment $t+1$' from 'the light is on at moment t' by a persistence axiom. However, what do we desire to derive from 'at t_1 the light should be on at moment t_2'?

1. Do we desire to infer that this obligation still holds at a later moment in time, i.e. 'at t_1+1 the light should be on at moment t_2,' or

2. Do we desire to infer that the obligation also holds for later moments, e.g.'at t_1 the light should be on at moment t_2+1?'

Moreover, what do we expect if we have an obligation that refers to an event instead of a fluent, such as 'at t_1 you should turn the light on at moment t_2?' In this paper we assume that only the first is derived, and it is derived for fluents as well as events, because this derivation is closest in spirit to the non-deontic derivation.[1] The persistence axioms for the deontic operators formalize the intuition that a deontic operator is only defeated by some explicit intervention.

2.1 Basic entities

The basic entities of Grünwald's first-order theories are fluents (properties of the world that persist), dependent fluents (that do not persist), events and (nonnegative integer) timepoints. He uses a many sorted first-order language $\mathcal{L}(\mathbf{D},\mathbf{E},\mathbf{F})$ that depends on the three sets $\mathbf{D}$, $\mathbf{E}$ and $\mathbf{F}$. It contains five sorts: booleans (with constants $\mathbf{B}=\{\text{TRUE},\text{FALSE}\}$), time points ($\mathbf{N}_0=\{0,1,2,\ldots\}$), events ($\mathbf{E}$), fluents ($\mathbf{F}$) and dependent fluents ($\mathbf{D}$). We write $\models$ for first-order entailment.

We extend this language in two ways. First we introduce finite sets of:

institutions, which we write as $\mathbf{I}=\{i_1,\ldots,i_n\}$, and

agents normally including an agent ϵ for external interventions (e.g. nature), which we write as $\mathbf{A}=\{\alpha_1,\ldots,\alpha_n,\epsilon\}$.

[1] There is also a technical problem when both derivations are accepted: blocking of the persistence rule becomes hard. For example, the obligation $O^{t+1}x^{t'+1}$ can be derived from $O^t x^{t'}$ in two ways, via $O^t x^{t'+1}$ and via $O^{t+1}x^{t'}$. Thus, if we block the derivation of $O^{t+1}x^{t'}$, then $O^{t+1}x^{t'+1}$ is still derived. Obviously, blocking the derivation of $O^{t+10}x^{t'+10}$ via $O^{t+5}x^{t'+5}$ is again much harder.

Second, we call the set $\mathbf{D} \cup \mathbf{E} \cup \mathbf{F}$ the brute facts, and we extend the set of brute facts with functions for deontic facts (not nested) and institutional facts (nested) as follows:

Deontic facts are new constants $O^t_\alpha x^{t'}$, $P^t_\alpha x^{t'}$, $F^t_\alpha x^{t'}$ and $W^t_\alpha x^{t'}$, for each agent $\alpha \in \mathbf{A}$, time points $t, t' \in \mathbf{N}_0$, and constant x of $\mathbf{D}$, $\mathbf{E}$ and $\mathbf{F}$, to be read as 'for agent α it is obligatory / explicitly permitted / forbidden / waived at moment t that x holds at moment t_x.' We write $O^t x^{t'}$ for $O^t_\epsilon x^{t'}$, etc.

Institutional facts are deontic facts together with new constants $Pow^t_\alpha \phi$ with ϕ an institutional fact.

The distinction between empowerment and permission is that the first is a dynamic and the latter a static notion: respectively a precondition for changing deontic states and a precondition for non-violation. For example, a priest may be empowered to marry two people in the sense that going through the ceremony counts as being married, but he may not be permitted to do so in the sense that he is violating a prohibition when he is going through the ceremony.

As discussed above, we assume that all deontic and institutional facts are fluents; but just like all other fluents they may be dependent or not, i.e. they persist or not. Dependent fluents are necessary to formalize the indirect effects of actions [Grü98], a discussion on this topic is outside the scope of this paper. We do not assume any relation between a brute fact and its deontic and institutional facts: if x is an independent fluent then Ox may still be a dependent fluent, and vice versa. The extended sets of constants are written as $\mathbf{D}^n$, $\mathbf{E}$ and $\mathbf{F}^n$, where n stands for normative. Usually we define predicates for $\mathbf{D}^n \cup \mathbf{E} \cup \mathbf{F}^n$ and we call their combined sort 'observables'. Note that these additional entities do not introduce any additional complexity; it is just more of the same.

The new constants correspond to modal logic expressions 'proposition p is obligatory' Op, 'p is explicitly permitted' Pp, 'p is forbidden' $Fp \equiv O\neg p$ and 'p is waived' $Wp \equiv P\neg p$.[2] We say that p is unnormed when all four modal sentences are false ($\neg Op \wedge \neg Pp \wedge \neg Fp \wedge \neg Wp$). Note that the main distinction between modal logic and our logic is that in modal logic we can

[2]Moreover, in modal logic sometimes 'p is weakly permitted' $P^- p \equiv \neg Fp \equiv \neg O \neg p$ is defined too. We do not consider weak permissions in this paper, because they do not increase the expressive power of the logic.

also encode principles such as the much debated $O(p \wedge q) \supset (Op \wedge Oq)$. Moreover, in some modal deontic logics nested operators are allowed. We do not incorporate these extensions because it sometimes they can be encoded in a different way (see below) and in general they do not seem to be necessary in computer science applications. Moreover, this lack of expressive power is the main reason that our first-order logic is more efficient than modal logic.

Grünwald's logic contains the following functions and predicates. We introduce the agent parameter of Do and we add the $Declare$ and Ctd predicates.

Ho : $Ho(x^t)$ denotes that observable x holds at time t;

Do : $Do_\alpha(x^t, b)$ denotes that an intervention takes place and the value of observable x at time t has been set to value b by an (unspecified) action of agent α; We say that there is a sufficient cause that x will (not) hold at timepoint t; We write $Do(x^t, b)$ for $Do_\epsilon(x^t, b)$;

$Declare$: $Declare^t_\alpha(x^{t'})$ denotes that if $Ho(Pow^t_\alpha(x^{t'}))$ then $Do_\alpha(x^{t'}, \text{TRUE})$

Ctd : $Ctd(O^t_\alpha(x^{t'}) \mid O^t_\alpha(y^t))$ denotes that if $Ho(O^t_\alpha(y^t)) \wedge \neg Ho(y^t)$ then $Do(O^t_\alpha(x^{t'}), \text{TRUE})$; analogously for prohibitions;

$=, <, +$: receive their usual interpretation;

Ab_1, Ab_2 : $Ab_1(e^t)$ and $Ab_2(f^t)$ both denote that something abnormal is the case concerning event e or fluent f at time t.

A distinction between Ho and Do is that the former associates two truth values with a constant (it is true or false) and the latter three (it can set the constant to true, to false, or neither). Another distinction is that if a fluent is true at a moment t, then we can derive in absence of interventions and abnormalities that the fluent was already true at $t-1$, an example of abduction. However, from setting this fluent to true at moment t we cannot derive anything for moment $t-1$. In the next section we show how this desired behavior is achieved by the persistence axioms.

Note that $Do_\alpha(e^t, \text{FALSE})$ formalizes that an agent refrains from action, i.e. prevents an event from occurring. This subject has got some attention in the philosophical literature on actions, but is not discussed further here.

It may seem restrictive that we cannot express for example that either fluent f_1 or fluent f_2 should hold, because we cannot express something like $O^t_\alpha(f_1 \vee f_2)$. However, this is not a restriction for a descriptive deontic logic, because we can simply define a different causal theory to describe this situation. For example, we can introduce a new variable f_3 and define its conditions by $Ho(f^t_3) \equiv Ho(f^t_1) \vee Ho(f^t_2)$, etc.

2.2 First-order causal theories

Grünwald's causal theories consist of two sets of axioms: CONS and EQ. CONS contains all axioms specific to the domain under consideration; EQ contains general axioms involving the persistence and the effects of interventions. In our deontic extension, we also introduce the equivalents of the modal deontic logic axioms $Op \supset Pp$ and $\neg(Op \wedge P\neg p)$ in CONS. When f^t stands for a deontic fact then t refers to the temporal index of the deontic operator, and when it refers to an empowerment, then t refers to the temporal index of the outer Pow operator.

Definition 1 (CausalDL) *A first-order causal theory with persistence, dependent fluents, concurrent events and surprises for a language* $\mathcal{L}(\mathbf{D}^n,\mathbf{E},\mathbf{F}^n)$ *is a tuple* $\langle \text{EQ}, \text{CONS} \rangle$ *where*

EQ *is a set of sentences containing the 'intervention axioms'*

$$\forall \alpha, x, t : Do_\alpha(x^t, \text{TRUE}) \supset Ho(x^t)$$

$$\forall \alpha, x, t : Do_\alpha(x^t, \text{FALSE}) \supset \neg Ho(x^t)$$

the persistence axioms

$$\forall f, t : [(t > 0) \wedge \neg\exists\alpha Do_\alpha(f^t, \text{TRUE}) \wedge \neg\exists\alpha Do_\alpha(f^t, \text{FALSE}) \wedge \neg Ab_2(f^t)] \supset [Ho(f^t) \equiv Ho(f^{t-1})]$$

$$\forall f, t : [(t > 0) \wedge \neg\exists\alpha Do_\alpha(f^t, \text{TRUE}) \wedge \neg\exists\alpha Do_\alpha(f^t, \text{FALSE}) \wedge Ab_2(f^t)] \supset \neg[Ho(f^t) \equiv Ho(f^{t-1})]$$

and the no-events axiom

$$\forall e, t : [\neg\exists\alpha Do_\alpha(e^t, \text{TRUE}) \wedge \neg\exists\alpha Do_\alpha(e^t, \text{FALSE})] \supset [Ho(e^t) \equiv Ab_1(e^t)]$$

CONS *is a set of sentences containing at least the relations between the normative operators*

$$\forall \alpha, x, t, t' : Ho(O^t_\alpha x^{t'}) \supset Ho(P^t_\alpha x^{t'})$$

$$\forall \alpha, x, t, t' : Ho(F^t_\alpha x^{t'}) \supset Ho(W^t_\alpha x^{t'})$$
$$\forall \alpha, x, t, t' : \neg(Ho(O^t_\alpha x^{t'}) \wedge Ho(W^t_\alpha x^{t'}))$$
$$\forall \alpha, x, t, t' : \neg(Ho(F^t_\alpha x^{t'}) \wedge Ho(P^t_\alpha x^{t'}))$$
$$\forall \alpha, x, t, t' : Do(O^t_\alpha x^{t'}, b) \supset Do(P^t_\alpha x^{t'}, b)$$
$$\forall \alpha, x, t, t' : Do(F^t_\alpha x^{t'}, b) \supset Do(W^t_\alpha x^{t'}, b)$$
$$\forall \alpha, x, t, t' : \neg(Do(O^t_\alpha x^{t'}, b) \wedge Do(W^t_\alpha x^{t'}, b))$$
$$\forall \alpha, x, t, t' : \neg(Do(F^t_\alpha x^{t'}, b) \wedge Do(P^t_\alpha x^{t'}, b))$$

$$\forall \alpha, x, y, t, t' : Declare^t_\alpha(Pow^t_\alpha(x^{t'})) \wedge Ho(Pow^t_\alpha(x^{t'})) \supset Do_\alpha(x^{t'}, \text{TRUE})$$

$$\forall \alpha, x, y, t, t' : [Ctd(O^t_\alpha(x^{t'}) \mid O^t_\alpha(y^t)) \wedge Ho(O^t_\alpha(y^t)) \wedge \neg Ho(y^t)] \supset Do(O^t_\alpha(x^{t'}), \text{TRUE})$$
$$\forall \alpha, x, y, t, t' : [Ctd(F^t_\alpha(x^{t'}) \mid O^t_\alpha(y^t)) \wedge Ho(O^t_\alpha(y^t)) \wedge \neg Ho(y^t)] \supset Do(F^t_\alpha(x^{t'}), \text{TRUE})$$
$$\forall \alpha, x, y, t, t' : [Ctd(O^t_\alpha(x^{t'}) \mid F^t_\alpha(y^t)) \wedge Ho(F^t_\alpha(y^t)) \wedge Ho(y^t)] \supset Do^t_\alpha(O(x^t), \text{TRUE})$$
$$\forall \alpha, x, y, t, t' : [Ctd(F^t_\alpha(x^{t'}) \mid F^t_\alpha(y^t)) \wedge Ho(F^t_\alpha(y^t)) \wedge Ho(y^t)] \supset Do^t_\alpha(F(x^t), \text{TRUE})$$

and uniqueness-of-names and domain closure axioms for the sets $\mathbf{B}$*,* $\mathbf{D}^n$*,* $\mathbf{E}$ *and* $\mathbf{F}^n$*. An uniqueness-of-names axiom for a finite set of constants* $X = \{X_1, \ldots, X_n\}$ *is the formula*

$$\wedge_{i \neq j} X_i \neq X_j$$

and a domain closure axiom for the set X *is the axiom*

$$\forall x : x = X_1 \vee \ldots \vee x = X_n$$

'Ought' does not imply 'is' or vice versa, which corresponds to the *non*-validity of the modal axioms $Op \supset p$, $p \supset Op$, $Pp \supset p$ and $p \supset Pp$. In our causal theories this corresponds to the *non*-validity of $Ho(O^t_\alpha x^{t'}) \supset Ho(x^{t'})$, $Ho(x^{t'}) \supset Ho(O^t_\alpha x^{t'})$, etc. The details of the stepwise minimization of the abnormality predicates is quite complicated (1. minimize (circumscribe) the Do predicate in CONS, keeping Ho, Ab_1 and Ab_2 fixed; 2. add the axioms in EQ to the theory resulting from 1.; 3. minimize the Ab-predicates in the theory resulting from 2.) and discussed extensively by Grünwald [Grü98]. We refer the interested reader to this literature, and just repeat the main definitions in Appendix A.

3 Normative causal theories

If there is a prohibition that today fences are forbidden, then normally tomorrow fences are forbidden too. However, as discussed above only the deontic operators persist, not the brute facts inside deontic operators. The following example illustrates how this can be modeled explicitly.

Example 1 *Let* $\mathbf{A} = \{\alpha, \epsilon\}$, $\mathbf{E} = \emptyset$, $\mathbf{F} = \{fence\}$, $\mathbf{D} = \emptyset$, *and let all deontic and institutional facts be in* F. *Moreover,* CONS *contains the following formula.*

$$\forall t, t' : [Ho(F_\alpha^t fence^{t'})] \supset Do(F_\alpha^t fence^{t'+1}, \textsc{true})$$

This persistence can be conditionalised to situations in which there is not already a fence (and likewise it can be conditionalised to normal situations).

$$\forall t, t' : [\neg Ho(fence^{t'}) \wedge Ho(F_\alpha^t fence^{t'})] \supset Do(F_\alpha^t fence^{t'+1}, \textsc{true})$$

With events we can only formalize that actions have to occur at a specified time point. However, usually something has to occur before a certain time point, the so-called deadline. The following example illustrates how this can be formalized with fluents. It also illustrates why deontic facts have two temporal references, one for the deontic operator and one for the brute facts.

Example 2 *Let* $\mathbf{A} = \{\alpha, \epsilon\}$, $\mathbf{E} = \{Deliver\}$, $\mathbf{F} = \{HasDelivered\}$, $\mathbf{D} = \emptyset$, *and let all deontic and institutional facts be in* F. *Moreover, let* CONS *contain the following formulas, in which* $t*$ *is the deadline.*

$$Do(O_\alpha^0 HasDelivered^{t*}, \textsc{true})$$
$$\forall t : Ho(Deliver^t) \supset Do_\alpha(HasDelivered^{t+1})$$
$$\neg Ho(HasDelivered^0)$$

It is not obligatory that goods are delivered at a specific moment in time, like in $Ho(O_\alpha^0 Deliver^{t*})$, *but they have to be delivered before the deadline* $t*$.

4 Ought-To-Do = Ought-To-Be a Sufficient Cause

When is it obligatory to do something? Somehow we have to take the distinction between observations and interventions into account. In a preference-based setting, there are two obvious choices [Pea93]. Assume, for example, that ϕ_1 is deontically preferred to ϕ_2 iff ϕ_1 implies a subset of the norm violations implied by ϕ_2. It is obligatory to $Do(x^t, \text{TRUE})$ when

1. $Do(x^t, \text{TRUE})$ is deontically preferred to $\neg Do(x^t, \text{TRUE})$, or
2. $Do(x^t, \text{TRUE})$ is deontically preferred to $Do(x^t, \text{FALSE})$.

In this section we sketch how these two choices can be implemented in causal deontic logic. An event or fluent is violated if the following holds.

$$V_\alpha(x^t) =_{\text{def}} (Ho(O^t_\alpha x^t) \wedge \neg Ho(x^t)) \vee (Ho(F^t_\alpha x^t) \wedge Ho(x^t))$$

The set of violations implied by a formula ϕ, written as $V(\phi)$, is now simply the set of violated fluents and events in the preferred models of the causal theory, where CONS is extended with ϕ.

Definition 2 *Let $V_\alpha(x^t)$ be a violated event or fluent, and let $V(\phi)$ be the set of violations preferentially implied by $\langle \text{EQ}, \text{CONS} \cup \{\phi\}\rangle$. A causal theory implies*

- $O_\alpha(x^t)$ *if* $V(Do_\alpha(x^t, \text{TRUE})) \subset V(\neg Do_\alpha(x^t, \text{TRUE}))$
- $P_\alpha(x^t)$ *if* $V(Do_\alpha(x^t, \text{TRUE})) \subseteq V(\neg Do_\alpha(x^t, \text{TRUE}))$
- $F_\alpha(x^t)$ *if* $V(\neg Do_\alpha(x^t, \text{TRUE})) \subset V(Do_\alpha(x^t, \text{TRUE}))$
- $W_\alpha(x^t)$ *if* $V(\neg Do_\alpha(x^t, \text{TRUE})) \subseteq V(Do_\alpha(x^t, \text{TRUE}))$
- $O'_\alpha(x^t)$ *if* $V(Do_\alpha(x^t, \text{TRUE})) \subset V(Do_\alpha(x^t, \text{FALSE}))$
- $P'_\alpha(x^t)$ *if* $V(Do_\alpha(x^t, \text{TRUE})) \subseteq V(Do_\alpha(x^t, \text{FALSE}))$
- $F'_\alpha(x^t)$ *if* $V(Do_\alpha(x^t, \text{FALSE})) \subset V(Do_\alpha(x^t, \text{TRUE}))$
- $W'_\alpha(x^t)$ *if* $V(Do_\alpha(x^t, \text{FALSE})) \subseteq V(Do_\alpha(x^t, \text{TRUE}))$

There are at least the following two complications which have be be dealt with in a more realistic approach.

- Some obligations are stronger than others. This can easily be incorporated in the definition by defining other preference relations between sets of violations.
- In the context of ought-to-do obligations, and more generally in the context of actions, usually the agent's abilities are taken into account. A simple but effective and often used approach is to partition the set of events and fluents into parameters and decision variables, called uncontrollable and controllable propositions in [Bou94].

We leave the further development of ought-to-do obligations for further research.

5 Related research

This distinction between observations (Ho) and interventions (Do) is a fundamental insight from decision theory [LR57], see for example the Do predicate in Jeffrey's logic of decision [Jef83]. The relation with causal reasoning has been studied in causal decision theory, and causal theories and networks have been studied extensively recently in the reasoning about uncertainty community in artificial intelligence [Pea96]. In deontic logic they have been discussed by Pearl [Pea93]. As Grünwald discusses, the causality used in his theories is strongly related to 'causes' in [Thi97], 'caused' in [Lin95, Lin96], and occlusion in [San94, GD96].

The Do operator discussed in this paper is radically different from well known modal operators like the $STIT$ operator [HB95]. Although both take as an argument a state description and they return an action, they do so in a completely different way. The Do operator is descriptive, in the sense that it assumes a subjective notion of causality which remains unexplained. Only the consequences of the action are described, i.e. its effect on the physical state of the world. The STIT operator, on the other hand, is objective in the sense that seeing-to-it-that is explained in terms of choices made by the agent. As argued extensively by Grünwald [Grü98], the causal theories are a model (in the non-logical sense) for domains of action and change. To put the distinction in the words of the devil's advocates: STIT theory cannot formalize the benchmark examples of nonmonotonic temporal reasoning, and causal theories do not attempt to explain what it means that an agent makes a choice (it only describes the consequences of this choice). It is an open problem whether the two can be unified.

We are not aware of any work on the combination of causal theories and deontic logic. Ong and Lee [OL94] have studied the use of abduction in bureaucratic rules, which they use to detect deontic dilemmas. They do not consider nonmonotonic temporal reasoning or causal rules, and they do not distinguish between situations where abduction is justified (observations) and situations where it is not (creations).

6 Conclusions

In this paper we argued that causal reasoning is relevant for deontic reasoning, and we introduced a descriptive temporal deontic logic, based on causal theories. Further research has to reveal the strengths and weaknesses of this approach.

We believe that this approach of using causal theories in deontic logic is well suited for applications of deontic logic in computer science such as the specification and validation of dynamic systems. Most importantly, the logic is able to formalize the benchmark example of nonmonotonic temporal reasoning. This is achieved by defining the 'correct' causal theories, i.e. the causal theories that derive the desired results. Moreover, causal models can be represented by causal graphs, related to the popular Bayesian networks, and several computational properties and algorithms have been studied [Pea96].

A Nonmonotonic reasoning

Grünwald determines the 'causal' models of causal theories in several steps.

1. minimize (circumscribe) the Do predicate in CONS, keeping Ho, Ab_1 and Ab_2 fixed (Definition 4 below);
2. add the axioms in EQ to the theory resulting from (1);
3. minimize the Ab-predicates in the theory resulting from (2) (Definition 5).

The models resulting from step (3) are the preferred causal models of the causal theory defined by CONS and EQ. Note that nonmonotonic inference is used in steps (1) and (3). We first give the (well-known) definition of circumscription of [Lif85].

Definition 3 (Section 9.8. of [Grü98]) *Let T be a first-order theory for some language $\mathcal{L}$, let P be a tuple of predicate constants for $\mathcal{L}$ and let Z be a tuple of function and/or predicate constants for $\mathcal{L}$ disjoint with P. For any two structures M_1 and M_2 for the language $\mathcal{L}$, we write $M_1 \leq^{P;Z} M_2$ if*

1. *$\mid M_1 \mid = \mid M_2 \mid$ (same universes), and*
2. *$M_1[[K]] = M_2[[K]]$ for every constant K not in P, Z, and*
3. *$M_1[[P_i]] \subseteq M_2[[P_i]]$ for every $P_i \in P$.*

A structure M is a model of $Circum(T; P; Z)$ iff M is minimal in the class of T with respect to $<^{P;Z}$.

In the following definition we write $Circum(\text{CONS}, Do)$ for the circumscription of Do in CONS with all other functions and predicates kept fixed.

Definition 4 (Def. 9.16 of [Grü98]) *A structure M for the language $\mathcal{L}$ is a model for the first-order causal theory T, written as $M \models_c T$, iff*

1. *$M \models \text{EQ} \wedge Circum(\text{CONS}; Do)$, and*
2. *Time-points in $\mathcal{L}$ are interpreted as the integers; '+' and '<' are interpreted accordingly.*

We write $\mid M \mid_X$ for the universe of a sort X, and we assume without loss of generality that $\mid M \mid_b = \mathbf{B}$, $\mid M \mid_d = \mathbf{D}^n$, $\mid M \mid_e = \mathbf{E}^n$, $\mid M \mid_f = \mathbf{F}^n$, and that M interprets all elements in $\mathbf{B}$, $\mathbf{D}^n$, $\mathbf{E}$ and $\mathbf{F}^n$ as themselves.

Weak abnormality Ab_1 is defined over event-time pairs and strong abnormality Ab_2 is defined over fluent-time pairs. The definition below says that the models with the smallest interpretations of the strong abnormality Ab_2 should be preferred, and, among the remaining models, the models with the smallest interpretations of the weak abnormality predicate Ab_1.

Definition 5 (Def. 9.23 of [Grü98]) *A model M is a preferred causal model for causal theory T if*

1. *$M \models_c T$, and*
2. *There is no other $M' \models_c T$ with $M'[Ab_2] \subset M[Ab_2]$, and*
3. *There is no other $M' \models_c T$ with $M'[Ab_2] = M[Ab_2]$ and $M'[Ab_1] \subset M[Ab_1]$.*

Grünwald shows how these causal theories can formalize most benchmark problems in nonmonotonic temporal reasoning.

References

[Bou94] C. Boutilier. Toward a logic for qualitative decision theory. In *Proceedings of the KR'94*, pages 75–86, 1994.

[Dig99] F. Dignum. Autonomous agents and norms. *Artificial Intelligence and Law*, 7:69–79, 1999.

[DP94] A. Darwiche and J. Pearl. Symbolic causal networks. In *Proceedings of the AAAI94*, 1994.

[FS99] B.S. Firozabadhi and M. Sergot. Power and permission in security systems. In *Proceedings of the ?*, page ?, 1999.

[GD96] J. Gustafsson and P. Doherty. Embracing occlusion in specifying the indirect effects of actions. In *Proceedings of the KR'96*, 1996.

[Gef97] H. Geffner. Causality, constraints and the indirect effect of actions. In *Proceedings of the IJCAI'97*, pages 555–560, 1997.

[Grü98] P. Grünwald. *The Minimum Description Length Principle and Reasoning under Uncertainty.* PhD thesis, University of Amsterdam, 1998.

[HB95] J.F. Horty and N. Belnap. The deliberative stit: a study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, pages 583–644, 1995.

[HM86] S. Hanks and D. McDermott. Default reasoning, non-monotonic logics and the frame problem. In *Proceedings of the AAAI'86*, pages 328–333, 1986.

[Jef83] R. Jeffrey. *The Logic of Decision.* University of Chicago Press, 2nd edition, 1983.

[Kau86] H.A. Kautz. The logic of persistence. In *Proceedings of the AAAI'86*, pages 401–405, 1986.

[KH99] C. Krogh and H. Herrestad. Hohfeld in cyberspace and other applications of normative reasoning in agent technology. *Artificial Intelligence and Law*, 6:81–96, 1999.

[Lia98] C.-J. Liau. A logic for reasoning about action, preference and commitment. In H. Prade, editor, *Proceedings of the ECAI'98*, pages 552–556, 1998.

[Lif85] V. Lifschitz. Computing circumscription. In *Proceedings of the IJCAI'85*, pages 121–127, 1985.

[Lin95] F. Lin. Embracing causality in specifying the indirect effect of actions. In *Proceedings of the IJCAI'95*, 1995.

[Lin96] F. Lin. Embracing causality in specifying the indeterminate effect of actions. In *Proceedings of the AAAI'96*, 1996.

[LR57] R.D. Luce and H. Raiffa. *Games and Decisions.* John Wiley, New York, 1957.

[Mak86] D. Makinson. On the formal representation of right relations. *Journal of Philosophical Logic*, 15:403–425, 1986.

[OL94] K. Ong and R.M. Lee. Detecting deontic dilemmas in bureaucratic rules: a first-order implementation using abduction. In *Proceedings of the Second Workshop on Deontic Logic in Computer Science (Δeon'94)*, pages 252–280, Oslo, 1994.

[Pea93] J. Pearl. From conditional oughts to qualitative decision theory. In *Proceedings of the UAI'93*, pages 12–20, 1993.

[Pea96] J. Pearl. Causation, action and counterfactuals. In *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK'96)*, pages 51–73, 1996.

[San94] E.J. Sandewall. *Features and Fluents*. Oxford University Press, 1994.

[Thi97] M. Thielscher. Ramification and causality. *Artificial Intelligence*, 89:317–364, 1997.

[TT98] W. Thoen and Y.H. Tan. A logical model of transfer of obligations in trade contracts. *Accounting, Management and Information systems*, 8:23–38, 1998.

[vdTT99] L. van der Torre and Y. Tan. Rights, duties and commitments between agents. In *Proceedings of the IJCAI'99*, pages 1239–1244, 1999.