The Equational Approach to Contrary-to-duty Obligations

Dov M. Gabbay

Bar-Ilan University, Ramat-Gan, Israel
King's College London, London, UK
University of Luxembourg, Luxembourg

Abstract. We apply the equational approach to logic to define numerical equational semantics and consequence relations for contrary to duty obligations, thus avoiding some of the traditional known paradoxes in this area. We also discuss the connection with abstract argumentation theory. Makinson and Torre's input output logic and Governatori and Rotolo's logic of violation.

1 Methodological Orientation

This paper gives equational semantics to contrary to duty obligations (CTDs) and thus avoids some of the known CTD paradoxes. The paper's innovation is on three fronts.

1. Extend the equational approach from classical logic and from argumentation [1,2] to deontic modal logic and contrary to duty obligations [5].
2. Solve some of the known CTD paradoxes by providing numerical equational semantics and consequence relation to CTD obligation sets.
3. Have a better understanding of argumentation semantics.

Our starting point in this section is classical propositional logic, a quite familiar logic to all readers. We give it equational semantics and define equational consequence relation. This will explain the methodology and concepts behind our approach and prepare us to address CTD obligations. We then, in Section 2, present some theory and problems of CTD obligations and intuitively explain how we use equations to represent CTD sets.

Section 3 deals with technical definitions and discussions of the equational approach to CTD obligations, Section 4 compares with input output logic, Section 5 compares with the logic of violation and and we conclude in Section 6 with general discussion and future research.

Let us begin.

1.1 Discussion and Examples

Definition 1. *Classical propositional logic has the language of a set of atomic propositions Q (which we assume to be finite for our purposes) and the connectives $\neg$ and $\wedge$. A classical model is an assignment $h : Q \mapsto \{0, 1\}$. h can be extended to all wffs by the following clauses:*

C. Beierle and C. Meghini (Eds.): FoIKS 2014, LNCS 8367, pp. 1–61, 2014.
© Springer International Publishing Switzerland 2014

– $h(A \wedge B) = 1$ *iff* $h(A) = h(B) = 1$
– $h(\neg A) = 1 - h(A)$

The set of tautologies are all wffs A such that for all assignments $h, h(A) = 1$.

The other connectives can be defined as usual

$$a \to b = def.\ \neg(a \wedge \neg b)$$
$$a \vee b = \neg a \to b = \neg(\neg a \wedge \neg b)$$

Definition 2.

1. *A numerical conjunction is a binary function $\mu(x, y)$ from $[0,1]^2 \mapsto [0,1]$ satisfying the following conditions*
   (a) *μ is associative and commutative*
   $\mu(x, \mu(y, z)) = \mu(\mu(x, y), z)$
   $\mu(x, y) = \mu(y, x)$
   (b) $\mu(x, 1) = x$
   (c) $x < 1 \Rightarrow \mu(x, y) < 1$
   (d) $\mu(x, y) = 1 \Rightarrow x = y = 1$
   (e) $\mu(x, 0) = 0$
   (f) $\mu(x, y) = 0 \Rightarrow x = 0$ *or* $y = 0$
2. *We give two examples of a numerical conjunction*

$$\mathbf{n}(x, y) = \min(x, y)$$
$$\mathbf{m}(x, y) = xy$$

For more such functions see the Wikipedia entry on t-norms [9]. However, not all t-norms satisfy condition (f) above.

Definition 3.

1. *Given a numerical conjunction μ, we can define the following numerical (fuzzy) version of classical logic.*
   (a) *An assignment is any function $\mathbf{h}$ from wff into $[0, 1]$.*
   (b) *$\mathbf{h}$ can be extended to $\mathbf{h}_\mu$ defined for any formula by using μ by the following clauses:*
      – $\mathbf{h}_\mu(A \wedge B) = \mu(\mathbf{h}_\mu(A), \mathbf{h}_\mu(B))$
      – $\mathbf{h}_\mu(\neg A) = 1 - \mathbf{h}_\mu(A)$
2. *We call μ-tautologies all wffs A such that for all $\mathbf{h}$, $\mathbf{h}_\mu(A) = 1$.*

Remark 1. Note that on $\{0, 1\}$, $\mathbf{h}_\mu$ is the same as h. In other words, if we assign to the atoms value in $\{0, 1\}$, then $\mathbf{h}_\mu(A) \in \{0, 1\}$ for any A. This is why we also refer to μ as “semantics”.

The difference in such cases is in solving equations, and the values they give to the variables $0 < x < 1$.

Consider the equation arising from $(x \to x) \leftrightarrow \neg(x \to x)$. We want

$$\mathbf{h}_\mathbf{m}(x \to x) = \mathbf{h}_\mathbf{m}(\neg(x \to x))$$

We get

$$(1 - \mathbf{m}(x))\mathbf{m}(x) = [1 - \mathbf{m}(x) \cdot (1 - \mathbf{m}(x))]$$

or equivalently

$$\mathbf{m}(x)^2 - \mathbf{m}(x) + \tfrac{1}{2} = 0.$$

Which is the same as

$$(\mathbf{m}(x) - \tfrac{1}{2})^2 + \frac{1}{4} = 0.$$

There is no real numbers solution to this equation.

However, if we use the **n** semantics we get

$$\mathbf{h_n}(x \to x) = \mathbf{h_n}(\neg(x \to x))$$

or

$$\min(\mathbf{n}(x), (1 - \mathbf{n}(x)) = 1 - \min(\mathbf{n}(x), 1 - \mathbf{n}(x))$$

$\mathbf{n}(x) = \frac{1}{2}$ is a solution.

Note that if we allow **n** to give values to the atoms in $\{0, \frac{1}{2}, 1\}$, then all formulas A will continue to get values in $\{0, \frac{1}{2}, 1\}$. I.e. $\{0, \frac{1}{2}, 1\}$ is closed under the function **n**, and the function $\nu(x) = 1 - x$.

Also all equations with **n** can be solved in $\{0, \frac{1}{2}, 1\}$.

This is not the case for **m**. Consider for the example the the equation corresponding to $x \equiv x \wedge \ldots \wedge x$, $(n + 1$ times).

The equation is $x = x^{n+1}$. We have the solutions $x = 0, x = 1$ and all roots of unity of $x^n = 1$.

Definition 4. *Let I be a set of real numbers $\{0, 1\} \subseteq I \subseteq [0, 1]$. Let μ be a semantics. We say that I supports μ iff the following holds:*

1. *For any $x, y \in I$, $\mu(x, y)$ and $\nu(x) = 1 - x$ are also in I.*
2. *By a μ expression we mean the following*
   (a) x is a μ expression, for x atomic
   (b) If X and Y are μ expressions then so are $\nu(X) = (1 - X)$ and $\mu(X, Y)$
3. *We require that any equation of the form $E_1 = E_2$, where E_1 and E_2 are μ expressions has a solution in I, if it is at all solvable in the real numbers.*

Remark 2. Note that it may look like we are doing fuzzy logic, with numerical conjunctions instead of t-norms. It looks like we are taking the set of values $\{0, 1\} \subseteq I \subseteq [0, 1]$ and allowing for assignments $\mathbf{h}$ from the atoms into I and assuming that I is closed under the application of μ and $\nu(x)$. For $\mu = \mathbf{n}$, we do indeed get a three valued fuzzy logic with the following truth table, Figure 1.

Note that we get the same system only because our requirement for solving equations is also supported by $\{0, \frac{1}{2}, 1\}$ for **n**.

The case for **m** is different. The values we need are all solutions of all possible equations. It is not the case that we choose a set I of truth values and close under **m**, and ν.

It is the case of identifying the set of zeros of certain polynomials (the polynomials arising from equations). This is an algebraic geometry exercise.

A	B	$\neg A$	$A \wedge B$	$A \vee B$	$A \to B$
0	0	1	0	0	1
0	$\frac{1}{2}$	1	0	$\frac{1}{2}$	1
0	1	1	0	1	1
$\frac{1}{2}$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	1	1
1	0	0	0	1	0
1	$\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$
1	1	0	1	1	1

Fig. 1.

Remark 3. The equational approach allows us to model what is considered traditionally inconsistent theories, if we are prepared to go beyond $\{0, 1\}$ values. Consider the liar paradox $a \leftrightarrow \neg a$. The equation for this is (both for **m** for **n**) $a = 1 - a$ (we are writing 'a' for '$\mathbf{m}(a)$' or '$\mathbf{n}(a)$' f). This solves to $a = \frac{1}{2}$.

1.2 Theories and Equations

The next series of definitions will introduce the methodology involved in the equational point of view.

Definition 5

1. (a) *A classical equational theory has the form*

$$\Delta = \{A_i \leftrightarrow B_i \mid i = 1, 2, \ldots\}$$

where A_i, B_i are wffs.

(b) *A theory is called a B-theory*[1] *if it has the form*

$$x_i \leftrightarrow A_i$$

where x_i are atomic, and for each atom y there exists at most one i such that $y = x_i$.

2. (a) *A function* **f**: *wff* $\to [0, 1]$ *is an μ model of the theory if we have that* **f** *is a solution of the system of equations* $\mathbf{Eq}(\Delta)$.

$$\mathbf{h}_\mu(A_i) = \mathbf{h}_\mu(B_i), i = 1, 2, \ldots$$

(b) *Δ is μ consistent if it has an μ model*

[1] B for Brouwer, because we are going to use Brouwer's fixed point theorem to show that theories always have models.

3. *We say that a theory Δ μ semantically (equationally) implies a theory Γ if every solution of* $\mathbf{Eq}(\Delta)$ *is also a solution of* $\mathbf{Eq}(\Gamma)$.
   We write
   $$\Delta \vDash_\mu \Gamma.$$
   Let $\mathbb{K}$ be a family of functions from the set of wff to $[0,1]$. We say that $\Delta \vDash_{(\mu,\mathbb{K})} \Gamma$ if every μ solution $\mathbf{f}$ of $\mathbf{Eq}(\Delta)$ such that $\mathbf{f} \in \mathbb{K}$ is also an μ solution of $\mathbf{Eq}(\Gamma)$.
4. *We write*
   $$A \vDash_\mu B$$
   iff the theory $\top \leftrightarrow A$ semantically (equationally) implies $\top \leftrightarrow B$.
   Similarly we write $A \vDash_{(\mu,\mathbb{K})} B$. In other words, if for all suitable solutions $\mathbf{f}$, $\mathbf{f}(A) = 1$ implies $\mathbf{f}(B) = 1$.

Example 1.

1. Consider $A \wedge (A \to B)$ does it $\mathbf{m}$ imply B? The answer is yes.
   Assume $\mathbf{m}(A \wedge (A \to B)) = 1$ then $\mathbf{m}(A)(1 - \mathbf{m}(A)(1 - \mathbf{m}(B))) = 1$. Hence $\mathbf{m}(A) = 1$ and $\mathbf{m}(A)(1 - \mathbf{m}(B)) = 0$. So $\mathbf{m}(B) = 1$.
   We now check whether we always have that $\mathbf{m}(A \wedge (A \to B) \to B) = 1$.
   We calculate $\mathbf{m}(A \wedge (A \to B) \to B) = [1 - \mathbf{m}(A \wedge (A \to B))(1 - \mathbf{m}(B))]$.
   $$= [1 - \mathbf{m}(A)(1 - \mathbf{m}(A)(1 - \mathbf{m}(B))x(1 - \mathbf{m}(B))]$$
   Let $\mathbf{m}(A) = \mathbf{m}(B) = \frac{1}{2}$. we get
   $$= [1 - \tfrac{1}{2}(1 - \tfrac{1}{2} \times \tfrac{1}{2}) \cdot \tfrac{1}{2} = 1 - \frac{3}{16} = \frac{13}{16}.$$
   Thus the deduction theorem does not hold. We have
   $$A \wedge (A \to B) \vDash B$$
   but
   $$\nvDash A \wedge (A \to B) \to B.$$
2. (a) Note that the theory $\neg a \leftrightarrow a$ is not $(\{0,1\}, \mathbf{m})$ consistent while it is $(\{0, \frac{1}{2}, 1\}, \mathbf{m})$ consistent.
   (b) The theory $(x \to x) \leftrightarrow \neg(x \to x)$ is not $([0,1], \mathbf{m})$ consistent but it is $(\{0, \frac{1}{2}, 1\}, \mathbf{n})$ consistent, but not $(\{0,1\}, \mathbf{n})$ consistent.

Remark 4. We saw that the equation theory $x \wedge \neg x \leftrightarrow \neg(x \wedge \neg x)$ has no solutions (no $\mathbf{m}$-models) in $[0,1]$. Is there a way to restrict $\mathbf{m}$ theories so that we are assured of solutions? The answer is yes. We look at B-theories of the form $x_i \leftrightarrow E_i$ where x_i is atomic and for each x there exists at most one clause in the theory of the form $x \leftrightarrow E$. These we called B theories. Note that if $x = \top$, we can have several clauses for it. The reason is that we can combine
$$\top \leftrightarrow E_1$$
$$\top \leftrightarrow E_2$$

into

$$\top \leftrightarrow E_1 \wedge E_2.$$

The reason is that the first two equations require

$$\mathbf{m}(E_i) = \mathbf{m}(\top) = 1$$

which is the same as

$$\mathbf{m}(E_1 \wedge E_2) = \mathbf{m}(E_1) \cdot \mathbf{m}(E_2) = 1.$$

If x is atomic different from $\top$, this will not work because

$$x \leftrightarrow E_i$$

requires $\mathbf{m}(x) = \mathbf{m}(E_i)$ while $x \leftrightarrow E_1 \wedge E_2$ requires $\mathbf{m}(x) = \mathbf{m}(E_1)\mathbf{m}(E_2)$.

The above observation is important because logical axioms have the form $\top \leftrightarrow A$ and so we can take the conjunction of the axioms and that will be a theory in our new sense.

In fact, as long as our μ satisfies

$$\mu(A \wedge B) = 1 \Rightarrow \mu(A) = \mu(B) = 1$$

we are OK.

Theorem 1. *Let Δ be a B-theory of the form*

$$x_i \leftrightarrow E_i.$$

Then for any continuous μ, Δ has a $([0, 1], \mu)$ model.

Proof. Follows from Brouwer's fixed point theorem, because our equations have the form

$$\mathbf{f}(\boldsymbol{x}) = \mathbf{f}(\boldsymbol{E}(\boldsymbol{x}))$$

in $[0, 1]^n$ where $\boldsymbol{x} = (x_1, \ldots, x_n)$ and $\boldsymbol{E} = (E_1, \ldots, E_n)$.

Remark 5. If we look at B-theories, then no matter what μ we choose, such theories have μ-models in $[0, 1]$. We get that all theories are μ-consistent. A logic where everything is consistent is not that interesting.

It is interesting, therefore, to define classes of μ models according to some meaningful properties. For example the class of all $\{0, 1\}$ models. There are other classes of interest. The terminology we use is intended to parallel semantical concepts used and from argumentation theory.

Definition 6. *Let Δ be a B-theory. Let* $\mathbf{f}$ *be a μ-model of Δ. Let A be a wff.*

1. *We say $\mathbf{f}(A)$ is crisp (or decided) if $\mathbf{f}(A)$ is either 0 or 1. Otherwise we say $\mathbf{f}(A)$ is fuzzy or undecided.*
2. *(a)* $\mathbf{f}$ *is said to be crisp if $\mathbf{f}(A)$ is crisp for all A.*

(b) We say that $\mathbf{f} \leq \mathbf{g}$*, if for all* A*, if* $\mathbf{f}(A) = 1$ *then* $\mathbf{g}(A) = 1$*, and if* $\mathbf{f}(A) = 0$ *then* $\mathbf{g}(A) = 0$.

We say $\mathbf{f} < \mathbf{g}$ *if* $\mathbf{f} \leq \mathbf{g}$ *and for some* $A, \mathbf{f}(A) \notin \{0, 1\}$ *but* $\mathbf{g}(A) \in \{0, 1\}$.

Note that the order relates to crisp values only.

3. *Define the* μ*-crisp (or* μ*-stable) semantics for* Δ *to be the set of all crisp* μ*-model of* Δ.
4. *Define the* μ*-grounded semantics for* Δ *to be the set of all* μ*-models* $\mathbf{f}$ *of* Δ *such that there is no* μ*-model* $\mathbf{g}$ *of* Δ *such that* $\mathbf{g} < \mathbf{f}$.
5. *Define the* μ*-preferred semantics of* Δ *to be the set of all* μ*-models* $\mathbf{f}$ *of* Δ *such that there is no* μ*-model* $\mathbf{g}$ *of* Δ *with* $\mathbf{f} < \mathbf{g}$.
6. *If* $\mathbb{K}$ *is a set of* μ *models, we therefore have the notion of* $\Delta \vDash_{\mathbb{K}} \Gamma$ *for two theories* Δ *and* Γ.

1.3 Generating B-theories

Definition 7. *Let* S *be a finite set of atoms and let* R_a *and* R_s *be two binary relations on* S*. We use* $\mathcal{A} = (S, R_a, R_s)$ *to generate a* B*-theory which we call the argumentation network theory generated on* S *from the attack relation* R_a *and the support relation* R_s.

For any $x \in S$*, let* $y_1, \ldots, y_m$ *be all the elements* y *of* S *such that* yR_ax *and let* $z_1, \ldots, z_n$ *be all the elements* z *of* S *such that* xR_sz *(of course* m, n *depend on* x*). Write the theory* $\Delta_{\mathcal{A}}$.

$$\{x \leftrightarrow \bigwedge z_j \wedge \bigwedge \neg y_i \mid x \in S\}$$

We understand the empty conjunction as $\top$.

These generate equations

$$x = \min(z_j, 1 - y_i)$$

using the $\mathbf{n}$ *function or*

$$x = (\Pi_j z_j)(\Pi_i(1 - y_i))$$

using the $\mathbf{m}$ *function.*

Remark 6.

1. If we look at a system with attacks only of the form $\mathcal{A} = (S, R_a)$ and consider the $\mathbf{n}(\min)$ equational approach for $[0, 1]$ then $\mathbf{n}$ models of the corresponding B-theory $\Delta_{\mathcal{A}}$ correspond exactly to the complete extensions of (S, R_a). This was extensively investigated in [1,2]. The semantics defined in Definition 6, the stable, grounded an preferred $\mathbf{n}$-semantics correspond to the same named semantics in argumentation, when restricted to B-theories arising from argumentation.
   If we look at μ other than $\mathbf{n}$, example we look at $\mu = \mathbf{m}$, we get different semantics and extensions for argumentation networks. For example the network of Figure 2 has the $\mathbf{n}$ extensions $\{a = 1, b = 0\}$ and $\{a = b = \frac{1}{2}\}$
   while it has the unique $\mathbf{m}$ extension $\{a = 1, b = 0\}$.

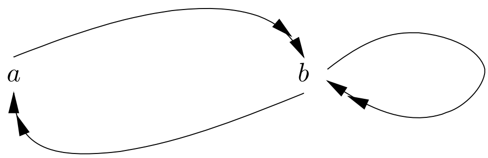


Fig. 2.

2. This correspondence suggests new concepts in the theory of abstract argumentation itself. Let $\Delta_{\mathcal{A}}, \Delta_{\mathcal{B}}$ be two B-theories arising from two abstract argumentation system $\mathcal{A} = (S, R_{\mathcal{A}})$ and $\mathcal{B} = (S, R_{\mathcal{B}})$ based on the same set S. Then the notion of $\Delta_{\mathcal{A}} \vDash_{\mathbb{K}} \Delta_{\mathcal{B}}$ as defined in Definition 5 suggest the following consequence relation for abstract argumentation theory.
   - $\mathcal{A} \vDash_{\mathbb{K}} \mathcal{B}$ iff any $\mathbb{K}$-extension ($\mathbb{K}$=complete, grounded, stable, preferred) of $\mathcal{A}$ is also a $\mathbb{K}$-extension of $\mathcal{B}$.

   So, for example, the network of Figure 3(a) semantically entails the network of Figure 3(b).

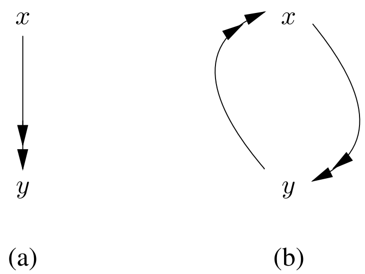


Fig. 3.

Remark 7. We can use the connection of equational B-theories with argumentation networks to export belief revision and belief merging from classical logic into argumentation. There has been considerable research into merging of argumentation networks. Classical belief merging offers a simple solution. We only hint here, the full study is elsewhere [10].

Let $\mathcal{A}_i = (S, R_i), i = 1, \ldots, n$, be the argumentation networks to be merged based on the same S. Let Δ_i be the corresponding equational theories with the corresponding semantics, based on **n**. Let $\mathbf{f}_i$ be respective models of Δ_i and let μ be a merging function, say $\mu = \mathbf{m}$.

Let $\mathbf{f} = \mu(\mathbf{f}_1, \ldots, \mathbf{f}_n)$. Then the set of all such $\mathbf{f}$s is the semantics for the merge result. Each such an **f** yields an extension.

Remark 8. The equational approach also allows us to generate more general abstract argumentation networks. The set S in (S, R_a) need not be a set of atoms. It can be a set of wffs.

Thus following Definition 7 and remark 6, we get the equations (for each A, B_j and where B_j are all the attackers of A:

$$\mathbf{f}(A) = \mu(\mathbf{f}(\neg B_1), \ldots,).$$

There may not be a solution.

2 Equational Modelling of Contrary to Duty Obligations

This section will use our μ-equational logic to model contrary to duty (CTD) sets of obligations. So far such modelling was done in deontic logic and there are difficulties involved. Major among them is the modelling of the Chisholm set [11].

We are going to use our equational semantics and consequence of Section 1 and view the set of contrary to duty obligations as a generator for an equational theory. This will give an acceptable paradox free semantics for contrary to duty sets.

We shall introduce our semantics in stages. We start with the special case of the generalised Chisholm set and motivate and offer a working semantical solution. Then we show that this solution does not work intuitively well for more general sets where there are loops. Then we indicate a slight mathematical improvement which does work. Then we also discuss a conceptual improvement.

The reader might ask why not introduce the mathematical solution which works right from the start? The answer is that we do not do this for reasons of conceptual motivation, so we do not appear to be pulling a rabbit out of a hat!

We need first to introduce the contrary to duty language and its modelling problems.

2.1 Contrary to Duty Obligations

Consider a semi-formal language with atomic variables $Q = \{p, q, r, \ldots\}$ the connective $\rightarrow$ and the unary operator $\bigcirc$. We can write statements like

1. $\bigcirc\neg$ fence
   You should not have a fence
2. fence $\rightarrow \bigcirc$ whitefence
   If you do have a fence it should be white.
3. Fact: fence

We consider a generalised Chisholm set of contrary to duty obligations (CTD) of the form

$$Oq_0$$

and for $i = 0, \ldots, n$ we have the CTD is

$$q_i \rightarrow Oq_{i+1}$$
$$\neg q_i \rightarrow O\neg q_{i+1}$$

and the facts $\pm q_j$ for some $j \in J \subseteq \{0, 1, \ldots, n+1\}$. Note that for the case of $n = 1$ and fact $\neg q_0$ we have the Chisholm paradox.

2.2 Standard Deontic Logic and Its Problems

A logic with modality $\Box$ is **KD** modality if we have the axioms

K0 All substitution instances of classical tautologies
K1 $\Box(p \wedge q) \equiv (\Box p \wedge \Box q)$
K2 $\vdash A \Rightarrow \vdash \Box A$
D $\neg\Box\bot$

It is complete for frames of the form (S, R, a) where $S \neq \oslash$ is a set of possible worlds, $a \in S, R \subseteq S \times S$ and $\forall x \exists y(xRy)$.

Standard Deontic Logic **SDL** is a **KD** modality O. We read $u \vDash Op$ as saying p holds in all ideal worlds relative to u, i.e. $\forall t(uRt \Rightarrow t \vDash p)$. So the set of ideal worlds relative to u is the set $I(u) = \{t \mid uRt\}$.

The **D** condition says $I(x) \neq \oslash$ for $x \in S$.

Following [8], let us quickly review some of the difficulties facing **SDL** in formalizing the Chisholm paradox.

The Chisholm Paradox

A. Consider the following statements:
  1. It ought to be that a certain man go to the assistance of his neighbour.
  2. It ought to be that if he does go he tell them he is coming.
  3. If he does not go then he ought not to tell them he is coming.
  4. He does not go.

  It is agreed that intuitively (1)–(4) of Chisholm set A are consistent and totally independent of each other. Therefore it is expected that their formal translation into logic **SDL** should retain these properties.

B. Let us semantically write the Chisholm set in semiformal English, where p and q as follows, p means HELP and q means TELL.
  1. Obligatory p.
  2. $p \rightarrow$ Obligatory q.
  3. $\neg p \rightarrow$ Obligatory $\neg q$.
  4. $\neg p$.

  Consider also the following:
  5. p.
  6. Obligatory q.
  7. Obligatory $\neg q$.

We intuitively accept that (1)–(4) of B are consistent and logically independent of each-other. Also we accept that (3) and (4) imply (7), and that (2) and (5) imply (6). Note that some authors would also intuitively expect to conclude (6) from (1) and (2).

Now suppose we offer a logical system **L** and a translation τ of (1), (2), (3), (4) of Chisholm into **L**.

For example **L** could be Standard Deontic Logic or **L** could be a modal logic with a dyadic modality $O(X/Y)$ (X is obligatory in the context of Y). We expect some coherence conditions to hold for the translation, as listed in Definition 8.

Definition 8. (Coherence conditions for representing contrary to duty obligations set in any logic)
We now list coherence conditions for the translation τ and for $\mathbf{L}$.
We expect the following to hold.

(a) *"Obligatory X" is translated the same way in (1), (2) and (3).*
Say τ(Obligatory X)=φ(X).

(b) *(2) and (3) are translated the same way, i.e., we translate the form:*
(23): $X \to$ Obligatory Y
to be $\psi(X, Y)$ and the translation does not depend on the fact that we have (4) $\neg p$ as opposed to (5) p.
Furthermore, we might, but not necessarily, expect $\psi(X/\top) = \varphi(X)$.

(c) *if X is translated as $\tau(X)$ then (4) is translated as $\neg\tau(X)$, the form (23) is translated as $\psi(\tau(X), \tau(Y))$ and (1) is translated as $\varphi(\tau(X))$.*

(d) *the translations of (1)–(4) remain independent in* $\mathbf{L}$ *and retain the connections that the translations of (2) and (5) imply the translation of (6), and the translations of (3) and (4) imply the translation of (7).*

(e) *the translated system maintains its properties under reasonable substitution in* $\mathbf{L}$.
The notion of reasonable substitution is a tricky one. Let us say for the time being that if we offer a solution for one paradox, say $\Pi_1(p, q, r, \ldots)$ and by substitution for $p, q, r, \ldots$ we can get another well known paradox Π_2, then we would like to have a solution for Π_2. This is a reasonable expectation from mathematical reasoning. We give a general solution to a general problem which yields specific solutions to specific problems which can be obtained from the general problem.

(f) *the translation is essentially linguistically uniform and can be done item by item in a uniform way depending on parameters derived from the entire database. To explain what we mean consider in classical logic the set*
(1) p
(2) $p \to q$.
To translate it into disjunctive normal form we need to know the number of atoms to be used. Item (1) is already in normal form in the language of $\{p\}$ but in the language of $\{p, q\}$ its normal form is $(p \wedge q) \vee (p \wedge \neg q)$. If we had another item
(3) r
then the normal form of p in the language of $\{p, q, r\}$ would be
$(p \wedge q \wedge r) \vee (p \wedge q \wedge \neg r) \vee (p \wedge \neg q \wedge r) \vee (p \wedge \neg q \wedge \neg r)$.
The moral of the story is that although the translation of (1) is uniform algorithmically, we need to know what other items are in the database to set some parameters for the algorithm.

Jones and Pörn, for example, examine in [8] possible translations of the Chisholm (1)–(4) into **SDL**. They make the following points:

(1) If we translate according to, what they call, option a:
  (1a) Op
  (2a) $O(p \to q)$

(3a) $\neg p \to O\neg q$
(4a) $\neg p$
then we do not have consistency, although we do have independence

(2) If we translate the Chisholm item (2) according to what they call option b:
(2b) $p \to Oq$
then we have consistency but not independence, since (4a) implies logically (2b).

(3) If (3a) is replaced by
(3b) $O(\neg p \to \neg q)$
then we get back consistency but lose independence, since (1a) implies (3b).

(4) Further, if we want (2) and (5) to imply (6), and (3) and (4) to imply (7) then we cannot use (3b) and (2a).

The translation of the Chisholm set is a "paradox" because known translations into Standard Deontic Logic (the logic with O only) are either inconsistent or dependent.

All the above statements together are logically independent and are consistent. Each statement is independent of all the others. If we want to embed the (model them) in some logic, we must preserve these properties and correctly get all intuitive inferences from them.

Remark 9. We remark here that the Chisholm paradox has a temporal dimension to it. The $\pm$tell comes before the $\pm$go. In symbols, the $\pm q$ is temporally before the $\pm p$. This is not addressed in the above discussion.

Consider a slight variation:

1. It ought to be that a certain man go to the assistance of his neighbour.
2. It ought to be that if he does not go he should write a letter of explanation and apology.
3. If he does go, then he ought not write a letter of explanation and apology.
4. He does not go.

Here $p =$ he does go and $q =$ he does not write a letter. Here q comes after p.

It therefore makes sense to supplement the Chisholm paradox set with a temporal clause as follows:

1. p comes temporally before q.

In the original Chisholm paradox the supplement would be:

1. Tell comes temporally before go.

2.3 The Equational Approach to CTD

We are now ready to offer equational semantics for CTD. Let us summarise the tools we have so far.

1. We have μ semantics for the language of classical logic.
2. Theories are sets of equivalences of the form $E_1 \leftrightarrow E_2$.
3. We associate equations with such equivalences.
4. Models are solutions to the equations.

5. Using models, we define consequence between theories.
6. Axioms have the for $\top \leftrightarrow E$
7. B-theories have the form $x \leftrightarrow E$, where x is atomic and E is unique to x.
8. We always have solutions for equations corresponding to B-theories.

Our strategy is therefore to associate a B-theory $\Delta(\mathbb{C})$ with any contrary to duty set $\mathbb{C}$ and examine the associated μ-equations for a suitable μ. This will provide semantics and consequence for the CTD sets and we will discuss how good this representation is.

The perceptive reader might ask, if Obligatory q is a modality, how come we hope to successfully model it in μ classical logic? Don't we need modal logic of it? This is a good question and we shall address it later. Of course modal logic can be translated into classical logic, so maybe the difficulties and paradoxes are "lost in translation". See Remark 15.

Definition 9. *1. Consider a language with atoms, the semi-formal $\rightarrow$ and $\neg$ and a semi-formal connective O.*
A contrary to duty expression has the form $x \rightarrow Oy$ where x and y are literals, i.e. either atoms q or negations of atoms $\neg q$, and where we also allow for x not to appear. We might write $\top \rightarrow Oy$ in this case, if it is convenient.

2. *Given a literal x and a set $\mathbb{C}$ of CTD expressions, then the immediate neighbourhood of x in $\mathbb{C}$. is the set $\mathbb{N}_x$ of all expressions from $\mathbb{C}$ of the form*

$$z \rightarrow Ox$$

or the form

$$x \rightarrow Oy.$$

3. *A set $\mathbb{F}$ of facts is just a set of literals.*
4. *A general CTD system is a pair $(\mathbb{C}, \mathbb{F})$*
5. *A Chisholm CTD set $\mathbb{CH}$ has the form*

$$x_i \rightarrow Ox_{i+1}$$
$$\neg x_i \rightarrow O\neg x_{i+1}$$
$$Ox_1$$

where $1 \leq i \leq m$ and x_i are literals (we understand that $\neg\neg x$ is x).

Example 2. Figure 4 shows a general CTD set

$$\mathbb{C} = \{a \rightarrow Ob, b \rightarrow O\neg a\}$$

Figure 5 shows a general Chisholm set. We added an auxiliary node x_0 as a starting point.

Figure 6 shows a general neighbourhood of a node x.

We employed in the figures the device of showing, whenever $x \rightarrow Oy$ is given, two arrows, $x \rightarrow y$ and $x \twoheadrightarrow \neg y$. The single arrow $x \rightarrow y$ means "from x go to y" and the double arrow $x \twoheadrightarrow \neg y$ means "from x do not go to $\neg y$".

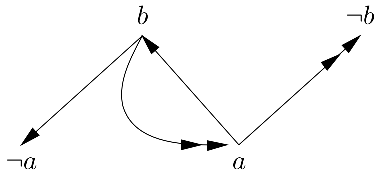


Fig. 4.

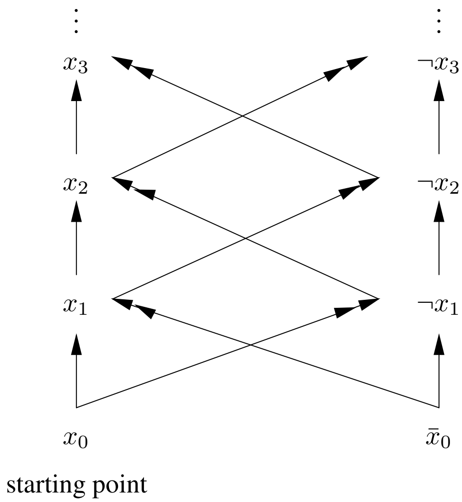


Fig. 5.

Remark 10. In Figures 4–6 we understand that an agent is at the starting point x_0 and he has to go along the arrows $\to$ to follow his obligations. He should not go along any double arrow, but if he does, new obligations (contrary to duty) appear.

This is a mathematical view of the CTD. The obligations have no temporal aspect to them but mathematically there is an obligation progression $(\pm x_0, \pm x_1, \pm x_2, \ldots)$.

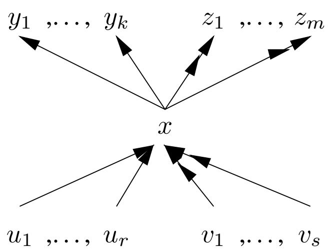


Fig. 6.

In the Chisholm example, the obligation progression is (± go, ±tell), while the practical temporal real life progression is (±tell, ±go). We are modelling the obligation progression.

To be absolutely clear about this we give another example where there is similar progression. Take any Hilbert axiom system for classical logic. The consequence relation $A \vdash B$ is timeless. It is a mathematical relation. But in practice to show $A \vdash B$ from the axioms, there is a progression of substitutions and uses of modus ponens. This is a mathematical progression of how we generate the consequence relations.

Remark 11. We want to associate equations with a given CTD set. This is essentially giving semantics to the set. To explain the methodology of what we are doing, let us take an example from the modal logic **S4**. This modal logic has wffs of the form $\Box q$. To give semantics for $\Box q$ we need to agree on a story for "□" which respects the logical theorems which "□" satisfies (completeness theorem). The following are possible successful stories about "□" for which there is completeness.

1. Interpret $\Box$ to mean provability in Peano arithmetic.
2. $\Box q$ means that q holds in all possible accessible situations (Kripke models).
3. $\Box$ means topological interior in a topological space.
4. $\Box$ means the English progressive tense:
   □ eat = "is eating"
5. $\Box$ means constructive provability.

For the case of CTD we need to adopt a story respecting the requirement we have on CTD.

Standard deontic logic **SDL** corresponds to the story that the meaning of OA in a world is that A holds in all accessible relative ideal worlds. It is a good story corresponding to the intuition that our obligations should take us to a better worlds. Unfortunately, there are difficulties with this story, as we have seen.

Our story is different. We imagine we are in states and our obligations tell us where we can and where we cannot go from our state. This is also intuitive. It is not descriptive as the ideal world story is, but it is operational , as real life is.

Thus in Figure 6 an agent at node x wants to say that he is a "good boy". So at x he says that he intends to go to one of $y_1, \ldots, y_k$ and that he did not come to x from $v_1, \ldots, v_k$, where the obligation was not to go to x.

Therefore the theory we suggest for node x is

$$x \leftrightarrow (\bigwedge_i y_i \wedge \bigwedge_j \neg v_j)$$

We thus motivated the following intuitive, but not final, definition.

Let $\mathbb{C}$ be a CTD set and for each x let $\mathbb{N}_x$ be its neighbourhood as in Figure 6.
We define the theory $\Delta(\mathbb{C})$ to be

$$\{x \leftrightarrow (\bigwedge_i y_i \wedge \bigwedge_j \neg v_j) \mid \text{ for all } \mathbb{N}_x\}. \quad (*1)$$

This definition is not final for technical reasons. We have literals "$\neg q$"and we do not want equivalences of the form $\neg q \leftrightarrow E$. So we introduce a new atom $\bar{q}$ to represent $\neg q$ with the theory $\bar{q} \leftrightarrow \neg q$.

So we take the next more convenient definition.

Definition 10.

1. *Let $\mathbb{C}$ be a CTD set using the atoms Q. Let $Q^* = Q \cup \{\bar{q} \mid q \in Q\}$, where $\bar{q}$ are new atoms.*
   Consider $\mathbb{C}^*$ gained from $\mathbb{C}$ by replacing any occurrence of $\neg q$ by $\bar{q}$, for $q \in Q$. Using this new convention Figure 5 becomes Figure 7.

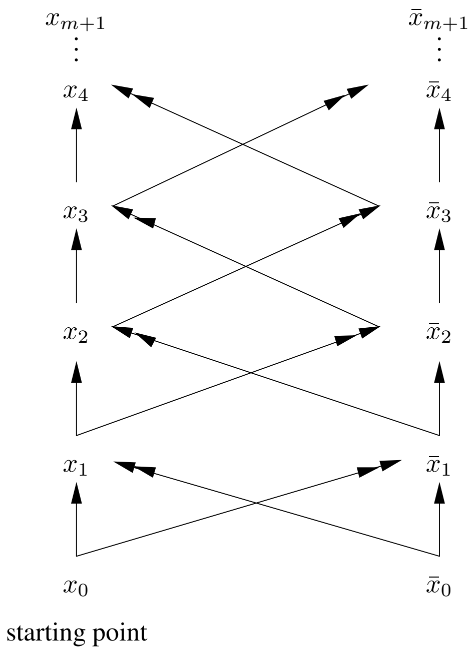


Fig. 7.

2. *The theory for the CTD set represented by Figure 7 is therefore*

$$
\begin{aligned}
&x_0 \leftrightarrow \top, \bar{x}_0 \leftrightarrow \bot\\
&x_0 \leftrightarrow x_1, \bar{x}_0 \leftrightarrow \bar{x}_1\\
&x_i \leftrightarrow x_{i+1} \wedge \bar{x}_{i-1}\\
&\bar{x}_i \leftrightarrow \bar{x}_{i+1} \wedge x_{i-1}\\
&\bar{x}_i \leftrightarrow \neg x_i\\
&x_{m+1} \leftrightarrow \bar{x}_m\\
&\bar{x}_{m+1} \leftrightarrow x_m\\
&\textit{for } 1 \leq i \leq m
\end{aligned}
$$

   The above is not a B-theory. The variable $\bar{x}_i$ has two clauses associated with it. (x_0 is OK because the second equation is $\top$). So is $\bar{x}_0$.
   It is convenient for us to view clause $\bar{x}_i = \neg x_i$ as an integrity constraint. So we have a B-theory with some additional integrity constraints.
   Note also that we regard all x_i and $\bar{x}_i$ as different atomic letters. If some of them are the same letter, i.e. $x_i = x_j$ then we regard that as having further integrity constraints of the form $x_i \leftrightarrow x_j$.
3. *The equations corresponding to this theory are*

$$
\begin{aligned}
&x_0 = 1, \bar{x}_0 = 0\\
&x_0 = x_1, \bar{x}_0 = \bar{x}_1\\
&x_i = \min(x_{i+1}, 1 - \bar{x}_{i-1})\\
&\bar{x}_i = \min(\bar{x}_{i+1}, 1 - x_{i-1})\\
&\bar{x}_i = 1 - x_i\\
&x_{m+1} = 1 - \bar{x}_m\\
&\bar{x}_{m+1} = 1 - x_m\\
&\textit{for } 1 \leq i \leq m
\end{aligned}
$$

   Remember we regard the additional equation

$$\bar{x}_i = 1 - x_i$$

   as an integrity constraint.
   Note also that we regard all x_i and $\bar{x}_i$ as different atomic letters. If some of them are the same letter, i.e. $x_i = x_j$ then we regard that as having further integrity constraints of the form $x_i \leftrightarrow x_j$. The rest of the equations have a solution by Brouwer's theorem. We look at these solutions and take only those which satisfy the integrity constraints. There may be none which satisfy the constraints, in which case the system overall has no solution!
4. *The dependency of variables in the equations of Figure 7 is described by the relation $x \Rightarrow y$ reading (x depends on y), where*

$$x \Rightarrow y = \textit{def.}\ (x \to y) \vee (y \twoheadrightarrow x).$$

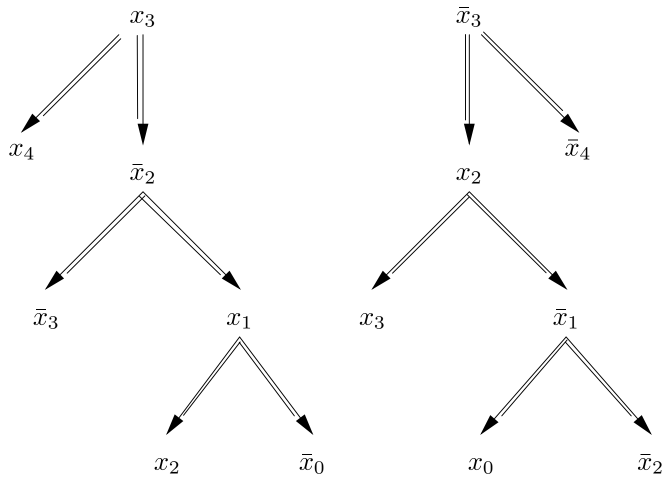


Fig. 8.

Figure 8 shows the variable dependency of the equations generated by Figure 7 up to level 3

Lemma 1.

1. *The equations associated with the Chisholm set of Figure 7 have the following unique solution, and this solution satisfies the integrity constraints:*

$$x_0 = 1, x_i = 1, \bar{x}_i = 0, \textit{ for } 0 \leq i \leq m + 1$$

2. *All the equations are independent.*

Proof.

1. By substitution we see the proposed solution is actually a solution. It is unique because $x_0 = 1$ and the variable dependency of the equations, as shown in Figure 8, is acyclic.
2. Follows from the fact that the variable dependency of the equations is acyclic. The variable x_i can depend only on the equations governing the variables below it in the dependency graph. Since it has the last equation in the tree, it cannot be derived from the equations below it.

Remark 12. We mentioned before that the theory (*1) and its equations above do not work for loops. Let us take the set $a \to \bigcirc\neg a$.

The graph for it, according to our current modelling would be Figure 9.

The equations for this figure would be

$$a = \min(1 - a, \bar{a})$$
$$a = 1 - \bar{a}$$

Fig. 9.

which reduces to

$$a = 1 - a$$
$$a = \tfrac{1}{2}$$

It does not have a consistent $\{0, 1\}$ solution.

We can fix the situation by generally including the integrity constraints $\bar{x} = 1 - x$ in the graph itself.

So Figure 9 becomes Figure 10, and the equations become

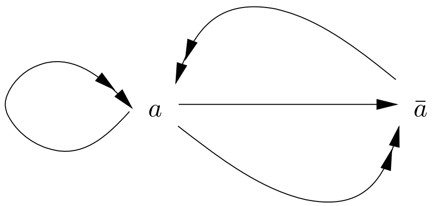


Fig. 10.

$$a = \min(\bar{a}, 1 - a, 1 - \bar{a})$$
$$\bar{a} = 1 - a$$

The two equations reduce to

$$a = \min(a, 1 - a)$$

which has the solution

$$a = 0, \bar{a} = 1$$

which fits our intuition.

Let us call this approach, (namely the approach where we do not view the equations $\bar{x} = 1 - x$ as integrity constraints but actually insert instead double arrow in the graph itself) the mathematical approach. What we have done here is to incorporate the integrity constraints $\bar{x} = 1 - x$ into the graph. Thus Figure 7 would become Figure 11, and the equations for the figure would become

$$\begin{aligned} x_i &= \min(x_{i+1}, 1 - \bar{x}_i, 1 - \bar{x}_{i-1}) \\ \bar{x}_i &= \min(\bar{x}_{i+1}, 1 - x_i, 1 - x_{i-1}) \\ x_0 &= 1, \bar{x}_0 = 0 \\ x_{m+1} &= \min(1 - \bar{x}_{m+1}, 1 - \bar{x}_m) \\ \bar{x}_{m+1} &= min(1 - x_{m+1}, 1 - x_m) \end{aligned}$$

for $1 \leq i \leq m$.

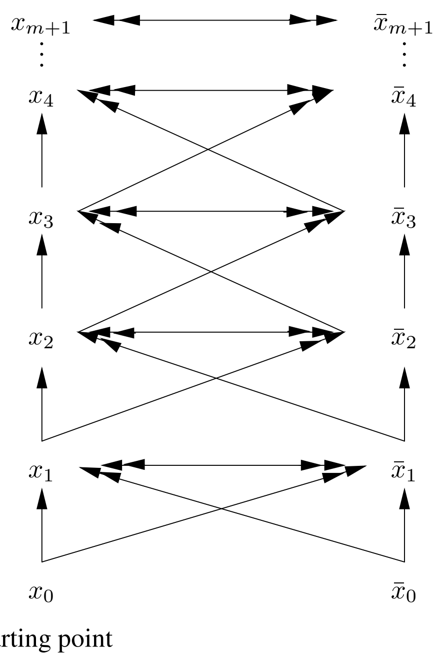


Fig. 11.

For the Chisholm set, we still get the same solution for these new equations, namely

$$\begin{aligned} x_0 = x_1 = \ldots x_{m+1} = 1 \\ \bar{x}_0 = \bar{x}_1 = \ldots = \bar{x}_{m+1} = 0 \end{aligned}$$

The discussion that follows in Definition 11 onwards applies equally to both graphs. We shall discuss this option in detail in Subsection 2.4.

The reader should note that we used here a mathematical trick. In Figure 11, there are two conceptually different double arrows. The double arrow $x_i \twoheadrightarrow x_{i+1}$ comes from an obligation $x_i \to \bigcirc x_{i+1}$, while the double arrows $x \twoheadrightarrow \bar{x}$ and $\bar{x} \twoheadrightarrow x$ come from logic (because $\bar{x} = \neg x$). We are just arbitratily mixing them in the graph!

Definition 11. *Consider Figure 7. Call this graph by $\mathbb{G}(m+1)$. We give some definitions which analyse this figure.*

First note that this figure can be defined analytically as a sequence of pairs

$$((x_0, \bar{x}_0), (x_1, \bar{x}_1), \ldots, (x_{m+1}, \bar{x}_{m+1})).$$

The relation $\to$ can be defined between nodes as the set of pairs $\{(x_i, x_{i+1})$ and $(\bar{x}_i, \bar{x}_{i+1})$ for $i = 0, 1, \ldots, m\}$. The relation $\twoheadrightarrow$ can be defined between nodes as the set of pairs $\{(x_i, \bar{x}_{i+1})$ and $(\bar{x}_i, x_{i+1})$ for $i = 0, 1, \ldots, m\}$. The starting point is a member of the first pair, in this case it is x_0, the left hand element of the first pair in the sequence, but we could have chosen $\bar{x}_0$ as the starting point.

1. *Let xRy be defined as $(x \to y) \vee (x \twoheadrightarrow y)$ and let R^* be the transitive and reflexive closure of R.*
2. *Let z be either x_i or $\bar{x}_i$. The truncation of $\mathbb{G}(m+1)$ at z is the subgraph of all points above z including z and $\bar{z}$ and all the arrow connections between them.*

   $$\mathbb{G}_z = \{y | zR^*y\} \cup \{\bar{z}\}$$

   We take z as the starting point of $\mathbf{G}(m+1)_z$. Note that $\mathbb{G}(m+1)_z$ is isomorphic to $\mathbb{G}(m+1-i)$. It is the same type of graph as $\mathbb{G}(m+1)$, only it starts at z. The corresponding equations for $\mathbb{G}_z$ will require $z = 1$.
3. *A path in the graph is a full sequence of points $(x_0, z_1, \ldots, z_{m+1})$ where z_i is $\bar{x}_i$ or x_i.*[2]
4. *A set of "facts" $\mathbb{F}$ in the graph is a set of nodes choosing at most exactly one of each pair $\{x_i, \bar{x}_i\}$.*
5. *A set of facts $\mathbb{F}$ restricts the possible paths by stipulating that the paths contain the nodes in the facts.*

Example 3. Consider Figure 7. The following is a path Π in the graph

$$\Pi = (x_0, x_1, x_2, x_3, \ldots, x_{m+1})$$

If we think in terms of an agent going along this path, then this agent committed two violations. Having gone to $\bar{x}_1$ instead of to x_1, he committed the first violation. From $\bar{x}_1$, the CTD says he should have gone to $\bar{x}_2$, but he went to x_2 instead. This is his second violation. After that he was OK.

Now look at the set of facts $= \{\bar{x}_1, x_2\}$. This allows for all paths starting with $(x_0, \bar{x}_1, x_2, \ldots)$. So our agent can still commit violations after x_2. We need more facts about his path.

[2] Note that the facts are sets of actual nodes. We can take the conjunction of the actual nodes as a formula faithfully representing the set of facts. Later on in this paper we will look at an arbitrary formula ϕ as generating the set of facts $\{y | y$ is either x_i or $\neg x_i$, unique for each i, such that $\phi | - y\}$.

According to this definition, $\phi = x_1 \vee x_2$, generates no facts. We will, however, find it convenient later in the paper, (in connection with solving the Miner's Paradox, Remark 20 below) to regard a disjunction as generating several possible sets of facts, one for each disjunct. See also Remark 19 below.

Suppose we add the fact $\bar{x}_4$.So our set is now $\mathbb{F} = \{\bar{x}_1, x_2, \bar{x}_4\}$.

We know now that the agent went from x_2 onto $\bar{x}_4$. The question is, did he pass through $\bar{x}_3$? If he goes to x_3, there is no violation and from there he goes to $\bar{x}_4$, and now there is violation.

If he goes to x_3, then the violation is immediate but when he goes from $\bar{x}_3$ to $\bar{x}_4$, there is no violation.

The above discussion is a story. We have to present it in terms of equations, if we want to give semantics to the facts.

Example 4. Let us examine what is the semantic meaning of facts. We have given semantic meaning to a Chisholm set $\mathbb{C}$ of contrary to duties; we constructed the graph, as in Figure 7 and from the graph we constructed the equations and we thus have equational semantics for $\mathbb{C}$.

We now ask what does a fact do semantically?

We know what it does in terms of our story about the agent. We described it in Example 3. What does a fact do to the graph? Let us take as an example the fact $\bar{x}_3$ added to the CTD set of Figure 7. What does it do? The answer is that it splits the figure into two figures, as shown in Figures 12 and 13.

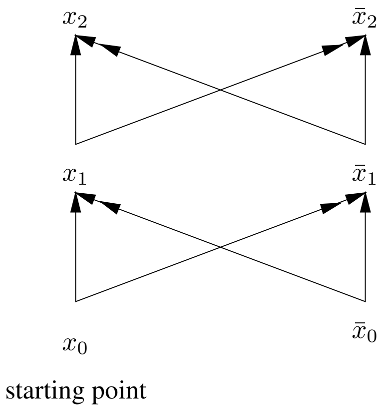


Fig. 12.

Note that Figure 13 is the truncation of Figure 7 at $\bar{x}_3$, and Figure 12 is the complement of this truncation.

Thus the semantical graphs and equations associated with $(\mathbb{C}, \{\bar{x}_3\})$ are the two figures, Figure 12 and Figure 13 and the equations they generate.

The "facts" operation is associative. Given another fact, say z it will be in one of the figures and so that figure will further split into two.

Definition 12. *Given a Chisholm system* $(\mathbb{C}, \mathbb{F})$ *as in Definition 9 we define its semantics in terms of graphs and equations. We associate with it with following system of*

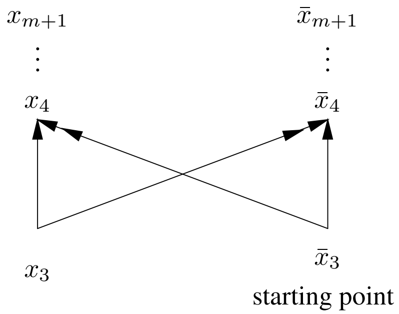


Fig. 13.

graphs (of the form of Figure 7) and these graphs will determine the equations, as in Definition 10.

The set $\mathbb{C}$ has a graph $\mathbb{G}(\mathbb{C})$. The set $\mathbb{F}$ can be ordered according to the relation R in the graph $\mathbb{G}(\mathbb{C})$ as defined in Definition 11. Let $(z_1, \ldots, z_k)$ be the ordering of $\mathbb{F}$.

We define by induction the following graphs:

1. *(a) Let $\mathbb{G}_k^+$ be $\mathbb{G}(\mathbb{C})_{z_k}$, (the truncation of $\mathbb{G}(\mathbb{C})$ at z_k). item Let $\mathbb{G}_k^-$ be $\mathbb{G}(\mathbb{C}) - \mathbb{G}_k^+$ (the remainder graph after deleting from it the top part $\mathbb{G}_k^+$).*
   (b) The point z_{k-1} is in the graph $\mathbb{G}_k^-$.
2. *Assume that for $z_i, 1 < i \leq k$ we have defined $\mathbb{G}_i^+$ and $\mathbb{G}_i^-$ and that $\mathbb{G}_i^+$ is the truncation of $\mathbb{G}_{i+1}^-$ at point z_i, and that $\mathbb{G}_i^- = \mathbb{G}_{i+1}^- - \mathbb{G}_i^+$. We also assume that z_{i-1} is in $\mathbb{G}_i^-$.*
   Let $\mathbb{G}_{i-1}^+ = (\mathbb{G}_i^-)_{z_{-1}}$, (i.e. the truncation of $\mathbb{G}_i^-$ at point z_{i-1}).
   Let $\mathbb{G}_{i-1}^- = \mathbb{G}_i^- - \mathbb{G}_{i-1}^+$.
3. *The sequence of graphs $\mathbb{G}, \mathbb{G}_1^-, \mathbb{G}_1^+, \mathbb{G}_2^+, \ldots, \mathbb{G}_k^+$ is the semantical object for $(\mathbb{C}, \mathbb{F})$. They generate equations which are the equational semantics for $(\mathbb{C}, \mathbb{F})$.*

Example 5. Consider a system $(\mathbb{C}, \mathbb{F})$ where $\mathbb{F}$ is a maximal path, i.e. $\mathbb{F}$ is the sequence $(z_1, \ldots, x_{m+1})$. The graph system for it will be as in Figure 14.

starting point	z_{m+1}	$\bar{z}_{m+1}$	graph $\mathbb{G}_{m+1}^+$
	$\vdots$	$\vdots$	$\vdots$
starting point	z_1	$\bar{z}_1$	graph $\mathbb{G}_1^+$
starting point	x_0	$\bar{x}_0$	graph $\mathbb{G}_1^-$

Fig. 14.

Remark 13. The nature of the set of facts $\mathbb{F}$ is best understood when the set $\mathbb{C}$ of Chisholm CTDs is represented as a sequence. Compare with Definition 12.

$\mathbb{C}$ has the graph $\mathbb{G}(\mathbb{C})$. The graph can be represented as a sequence

$$\mathbf{E} = ((x_0, \bar{x}_0), (x_1, \bar{x}_1), \ldots, (x_{m+1}, \bar{x}_{m+1}))$$

together with the starting point (x_0).

When we get a set of facts $\mathbb{F}$ and arrange it as a sequence $(z_1, \ldots, z_k)$ in accordance with the obligation progression, we can add x_0 to the sequence and look at $\mathbb{F}$ as

$$\mathbb{F} = (x_0, z_1, \ldots, z_k).$$

We also consider $(\mathbb{E}, \mathbb{F})$ as a pair, one the sequence $\mathbb{E}$ and the other as a multiple sequence of starting points. The graph $\mathbb{G}_i$ is no more and no less than the subsequence $\mathbb{E}_i$, beginning from the pair $(z_i, \bar{z}_i)$ up to the pair $(z_{i+1}, \bar{z}_{i+1})$ but *not* including $(z_{i+1}, \bar{z}_{i+1})$.

This way it is easy to see how $\mathbb{G}$ is the sum of all the $\mathbb{G}_i$, strung together in the current progression order. Furthermore, we can define the concept of "the fact z_j is in violation of the CTD of z_i", for $i < j$. To find out if there was such a violation, we solve the equations for

$$\mathbb{E}_i = ((z_i, \bar{z}_i), \ldots, (x_{m+1}, \bar{x}_{m+1}))$$

and if the equation solves with $z_j = 0$ then putting $z_j = 1$ is a violation.

Remark 14. Let us check whether our equational modelling of the Chisholm CTD set satisfies the conditions set out in Definition 8.

Consider Figure 15 (a) and (b):

(a) Obligatory x must be translated the same way throughout.
This holds because we use a variable x in a neighbourhood generated equation.

(b) The form $X \to OY$ must be translated uniformly no matter whether $X = q$ or $X = \neg q$.
This is is true of our model.

(c) This holds because "X" is translated as itself.

(d) The translation of the clauses must be all independent.
Indeed this holds by Lemma 1.
It is also true that (see Figure 15(a))

2. $p \to Oq$

and

5. p

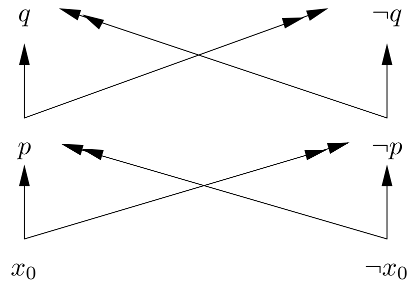
imply

6. Oq

This holds because (5) p is a fact. So this means that Figure 15(b) truncated at the point p.
The truncated figure is indeed what we would construct for Oq.
A symmetrical argument shows that (4) and (3) imply (7).

1. Op
2. $p \to Oq$
3. $\neg p \to Oq$
4. $\neg p$
5. p
6. Oq
7. $O\neg q$

(a)



starting point

(b)

Fig. 15.

(e) The system is required to be robust with respect to substitution.
This condition arose from criticism put forward in [7] against the solution to the Chisholm paradox offered in [8]. [8] relies on the fact that p, q are independent atoms.
The solution does not work when $q \vdash p$, e.g. substituting for "q" the wff "$r \wedge p$" (like $p =$ fence and $q =$ white fence).
In our case we use equations and if we substitute "$r \wedge p$" for "q" we get the equations

$$r \wedge p = 1 - p$$
$$p = r \wedge p$$

Although this type of equation is not guaranteed a solution, there is a solution in this case;$p = r = 1$.
If we add the fact $\neg p$, i.e. $1 - p = 1, p = 0$, (there is no fence) the equation solves to $\neg q = \neg p \vee \neg r$, which is also $= 1$ because of $\neg p$. So we have no problem with such substitution. In fact we have no problem with any substitution because the min function which we use always allows for solutions.

(f) The translation must be uniform and it to be done item by item.
Yes. Indeed, this is what we do!

Remark 15. We can now explain how classical logic can handle CTD, even though the CTD $x \to Oy$ involves a modality. The basic graph representation such as Figure 7 can be viewed as a set of possible worlds where the variables x and y act as nominals (i.e. atoms naming worlds by being true exactly at the world they name). x is a world, y is a world and $x \to y$ means y is ideal for x. $x \twoheadrightarrow \bar{y}$ means that $\bar{y}$ is sub ideal for x. Let $\Box_1$ be the modality for $\to$ and $\Box_2$ the modality for $\twoheadrightarrow$. Then we have a system with two disjoint modalities and we can define

$$OA \equiv \Box_1 A \wedge \Box_2 \neg A.$$

Now this looks familiar and comparable to [8], and especially to [12]. The perceptive reader might ask, if we are so close to modal logic, and in the modal logic formulation there are the paradoxes, why is it that we do not suffer from the paradoxes in the equational formulation?

The difference is because of how we interpret the facts! The equational approach spreads and inserts the facts into different worlds according to the obligation progression. Modal logic cannot do that because it evaluates formulas in single worlds. With equations, each variable is a nominal for a different world but is also is natural to substitute values to several variables at the same time!

Evaluating in several possible worlds at the same time in modal logic would solve the paradox but alas, this is not the way it is done.

Another difference is that in modal logic we can iterate modalities and write for example

$$O(x \to Oy).$$

We do not need that in Chisholm sets. This simplifies the semantics.

2.4 Looping CTDs

So far we modelled the Chisholm set only. Now we want to expand the applicability of the equational approach and deal with looping CTDs, as in the set in Figure 4. Let us proceed with a series of examples.

Example 6. Consider the CTD set of Figure 4. If we write the equations for this example we get

1. $a = \min(b, 1 - b)$
2. $b = \neg a$
3. $\neg b = 1 - a$

and the constants

4. $\neg b = 1 - b$
5. $\neg a = 1 - a.$

The only solution here is $a = b = \frac{1}{2}$. In argumentation and in classical logic terms this means the theory of Figure 4 is $\{0, 1\}$ inconsistent.

This is mathematically OK, but is this the correct intuition? Consider the set $\{b, \neg a\}$. The only reason this is not a solution is because we have $a \twoheadrightarrow \neg b$ and if $a = 0$, we get $\neg b = 1$ and so we cannot have $b = 1$.

However, we wrote $a \twoheadrightarrow \neg b$ because of the CTD $a \to Ob$, which required us to go from a to b (i.e $a \to b$) and in this case we put in the graph $a \twoheadrightarrow \neg b$ to stress "do not go to $\neg b$".

However, if $a = 0$, why say anything? We do not care in this case whether the agent goes from a to b!

Let us look again at Figure 6. We wrote the following equation for the node x

$$x = \min(u_i, 1 - v_j).$$

The rationale behind it was that we follow the rules, so we are going to u_i as our obligations say, and we came to x correctly, not from v_j, because $v_j \to O\neg x$ is required. Now if $v_j = 1$ (in the final solution) then the equation is correct. But if $v_j = 0$, then we do not care if we come to x from v_j, because $v_j \to O\neg x$ is not activated. So somehow we need to put into the equation that we care about v_j only when $v_j = 1$.

Remark 16. Let us develop the new approach mentioned in Example 6 and call it the soft approach. We shall compare it with the mathematical approach of Remark 12.

First we need a δ function as follows:

$$\delta(w) = \bot \text{ if } w = \bot$$

and

$$\delta(w) = \top \text{ if } w \neq \bot.$$

$\delta(w) = w$, if we are working in two valued $\{0, 1\}$ logic. Otherwise it is a projective function

$$\delta(0) = 0 \text{ and } \delta(w) = 1 \text{ for } w > 0.$$

We can now modify the equivalences (*1) (based on figure 6) as follows:

Let $_1, \ldots, v_s$ be as in Figure 6. Let $J, K \subseteq \{1, \ldots, s\}$ be such that $J \cap K = \varnothing$ and $J \cup K = \{1, \ldots, s\}$. Consider the expression

$$\varphi_{J,K} = \bigwedge_{j \in J} \delta(v_j) \wedge \bigwedge_{k \in K} \neg\delta(v_k).$$

This expression is different from 0 (or $\bot$), exactly when K is the set of all indices k for which $v_j = \bot$.

Replace (*1) by the following group of axioms for each pair J, K and for each x

$$x \wedge \varphi_{J,K} \leftrightarrow \varphi_{J,K} \wedge \bigwedge_{r} u_r \wedge \bigwedge_{j \in J} \neg v_j. \qquad (*2)$$

Basically what (*2) says is that the value of x should be equal to

$$\min\{u_r, 1 - v_j \text{ for those } j \text{ whose value is } \neq 0\}.$$

Note that this is an implicit definition for the solution of the equations. It is clear when said in words but looks more complicated when written mathematically. Solutions may not exist.

Example 7. Let us now look again at Figure 4.

The soft equations discussed in Remark 16 are

$$\begin{aligned}\delta(a) \wedge \bar{b} &= \delta(a)(1-a)\\ \delta(b) \wedge a &= \delta(b)\min(b, 1-b)\\ b &= \bar{a}\\ \bar{b} &= 1-b\\ \bar{a} &= 1-a.\end{aligned}$$

For these equations $\bar{a} = 1, \bar{b} = a = 0, b = 1$ is a solution.

Note that $\bar{a} = \bar{b} = 1$ and $a = b = 0$ is *not* a solution!

Let us now examine and discuss the mathematical approach alternative, the one mentioned in Remark 12. The first step we take is to convert Figure 4 into the right form for this alternative approach by adding double arrows between all x and $\bar{x}$. We get Figure 16.

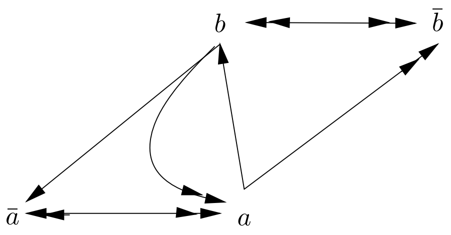


Fig. 16.

The equations are the following:

$$\begin{aligned}a &= \min(b, 1-\bar{a}, 1-b)\\ \bar{a} &= 1-a\\ b &= \min(\bar{a}, 1-\bar{b})\\ \bar{b} &= \min(1-a, 1-b).\end{aligned}$$

Let us check whether $a = \bar{b} = 0$ and $b = \bar{a} = 1$ is a solution. We get respectively by substitution

$$\begin{aligned}0 &= \min(1,0,0)\\ 1 &= 1-0\\ 1 &= \min(1, 1-0)\\ 0 &= min(1-0, 1-1).\end{aligned}$$

Indeed, we have a solution. Let us try the solution $\bar{b} = \bar{a} = 1$ and $a = b = 0$. Substitute in the equations and get

$$\begin{aligned} 0 &= \min(0, 0, 1) \\ 1 &= 1 - 0 \\ 0 &= \min(1, 1 - 1) \\ 1 &= \min(1 - 0, 1 - 0). \end{aligned}$$

Again we have a solution.

This solution also makes sense. Note that this is not a solution of the previous soft approach!

We need to look at more examples to decide what approach to take, and which final formal definition to give.

Example 8. Consider the following two CTD sets, put forward by two separate security advisors D and F.

D1: you should have a dog
Od
D2: If you do not have a dog, you should have a fence
$\neg d \to Of$
D3: If you have a dog you should not have a fence
$d \to O\neg f$
F1: You should have a fence
Of
F2: If you do not have a fence you should have a dog
$\neg f \to Od$
F3: If you do have a fence you should not have a dog.
$f \to O\neg d$

If we put both sets together we have a problem. They do not agree, i.e. {D1, D2, D3, F1, F2, F3}. However, we can put together both D1, D2 and F1, F2. They do agree, and we can have both a dog and a fence.

The mathematical equational modelling of D1 and D2 also models D3, i.e. D1, D2 $\vDash$ D3 and similarly F1, F2 $\vDash$ F3. So according to this modelling {D1, D2, F1, F2} cannot be consistently together. Let us check this point. Consider Figure 17

The equations for Figure 17 are:

$$\begin{aligned} x_0 &= 1 \\ x_0 &= d \\ \bar{x}_0 &= 1 - x_0 \\ d &= 1 - \bar{d} \\ \bar{d} &= \min(1 - d, 1 - x_0) \\ \bar{d} &= f \\ f &= 1 - \bar{f} \\ \bar{f} &= \min(1 - f, 1 - \bar{d}) \end{aligned}$$

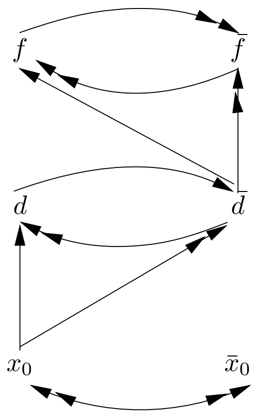


Fig. 17.

The only solution is

$$x_0 = d = \bar{f} = 1$$
$$\bar{x}_0 = \bar{d} = f = 0.$$

The important point is that $\bar{f} = 1$, i.e. no fence.

Thus D1,D2 $\vdash \bar{f}$.

By complete symmetry beget that F1,F2 $\vdash \bar{d}$. Thus we cannot have according to the mathematical approach that having both a dog and a fence is consistent with $\{$D1,D2, F1,F2$\}$.

Let us look now at the soft approach. Consider Figure 18

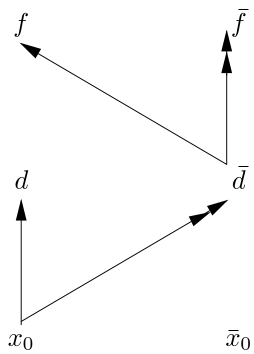


Fig. 18.

The soft equations for Figure 18 are:

$$x_0 = 1$$
$$x_0 = d$$
$$\min(x_0, \bar{d}) = \min(x_0, 1 - x_0)$$
$$\min(\bar{d}, \bar{f}) = \min(\bar{d}, 1 - \bar{d})$$

There are two solutions

$$x_0 = 1, d = 1, \bar{d} = 0, \bar{f} = 1, f = 0$$

and

$$x_0 = 1, d = 1, \bar{d} = 0, \bar{f} = 0, f = 1.$$

The conceptual point is that since $\bar{d} = 0$, we say nothing about $\bar{f}$.

Now similar symmetrical solution is available for {F1,F2} Since D1,D2 allow for $f = 1$ and F1,F2 allow for $d = 1$, they are consistent together. In view of this example we should adopt the soft approach.

Remark 17. Continuing with the previous Example 8, let us see what happens if we put together in the same CTD set the clauses {D1,D2,E1,E2} and draw the graph for them all together, in contrast to what we did before, where we were looking at two separate theories and seeking a joint solution. If we do put them together, we get the graph in Figure 19.

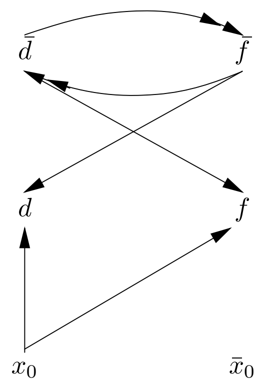


Fig. 19.

If we use the mathematical equations, there will be no solution. If we use the soft approach equations, we get a unique solution

$$d = f = 1, \bar{d} = \bar{f} = 0$$

The reason for the difference, I will stress again, is in the way we write the equations for $\bar{d}$ and $\bar{f}$. In the mathematical approach we write

$$\begin{aligned} \bar{d} &= \min(f, 1 - \bar{f}) \\ \bar{f} &= \min(d, 1 - \bar{d}) \\ \bar{d} &= 1 - d \\ \bar{f} &= 1 - f \end{aligned}$$

In the soft approach we write

$$\begin{aligned} \min(\bar{d}, \bar{f}) &= \min(f, \bar{d}, 1 - \bar{f}) \\ \min(\bar{f}, \bar{d}) &= \min(d, \bar{f}, 1 - d) \end{aligned}$$

This example also shows how to address a general CTD set, where several single arrows can come out of a node (in our case x_0). The equations for x_0 in our example are:

$$\begin{aligned} x_0 &= 1 \\ x_0 &= min(f, d) \end{aligned}$$

which forces $d = f = 1$. We will check how to generalise these ideas in the next section.

2.5 Methodological Discussion

Following the discussions in the previous sections, we are now ready to give general definitions for the equational approach to general CTD sets. However, before we do that we would like to have a methodological discussion. We aleady have semantics for CTD. It is the soft equations option discussed in the previous subsection. So all we need to do now is to define the notion of a general CTD set (probably just a set of clauses of the form $\pm x \to O \pm y$) and apply the soft equational semantics to it. This will give us a consequence relation and a consistency notion for CTD sets and the next step is to find proof theory for this consequence and prove a completeness theorem.

We need to ask, however, to what extent is the soft semantics going to be intuitive and compatible with our perception of how to deal with conflicting CTD sets? So let us have some discussion about what is intuitive first, before we start with the technical definitions in the next section. Several examples will help.

Example 9. Consider the following CTD set:

1. You should not have a dog
   $O\neg d$
2. If you have a dog you must keep it
   $d \to Od$
3. d: you have a dog

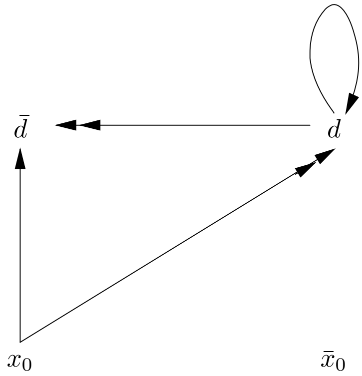


Fig. 20.

Here we have a problem. Is (1), (2), (3) a consistent set? In **SDL** we can derive from (2) and (3) $O\neg d$ and get a contradiction $Od \wedge O\neg d$.

However, in our semantics we produce a graph and write equations and if we have solutions, then the set is consistent. Let us do this.

The original graph for clauses (1)–(2) is Figure 20. This graph generates equations.

The fact d splits the graph and we get the two graphs in Figures 21 and 22.

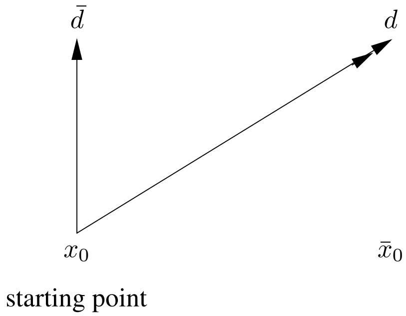


Fig. 21.

The solution of the soft equations for the original graph (without the fact d) is $x_0 = \bar{d} = 1, d = 0$.

The solution for the two split graphs, after the fact d gives $d = 1$ for Figure 21 and $\bar{d} = 0$ for Figure 22.

There is no mathematical contradiction here. We can identify a violation from the graphs. However we may say there is something unintuitive, as the CTD proposal for a remedy for the violation $O\neg d$, namely $d \rightarrow Od$ violates the original obligation $O\neg d$, and actually perpetuates this violation. This we see on the syntactical level. No problem in the semantics.

Fig. 22.

We can explain and say that since the fact d violates $O\neg d$ then a new situation has arisen and $O\neg d$ is not "inherited" across a CTD. In fact, in the case of a dog it even makes sense. We should not have a dog but if we violate the obligation and get it, then we must be responsible for it and keep it.

The next example is more awkward to explain.

Example 10. This example is slightly more problematic. Consider the following.

1. You should not have a dog
   $O\neg d$
2. you should not have a fence
   $O\neg f$
3. If you do have a dog you should have a fence
   $d \to Of$

The graph for (1)–(3) is Figure 23.

The solution is $\bar{d} = \bar{f} = 1, d = f = 1$.

Let us add the new fact

4. d: You have a dog

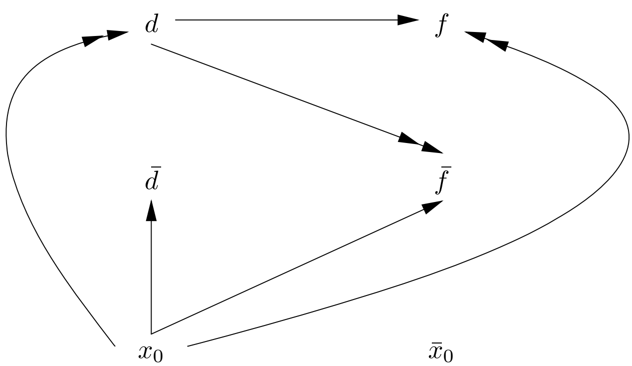


Fig. 23.

The graph of Figure 23 splits into two graphs, Figure 24 and 25.

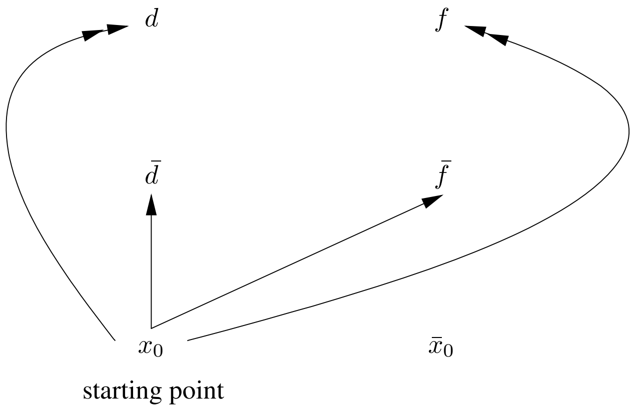


Fig. 24.

The equations for Figure 24 solve to $\bar{d} = \bar{f} = 1, d = f = 0$. The equations for Figure 25 solve to $d = f = 1, \bar{d} = \bar{f} = 0$.

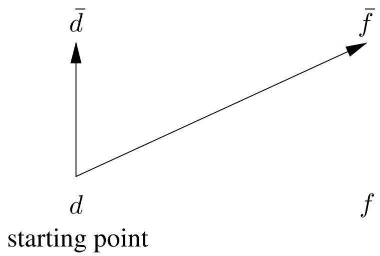


Fig. 25.

There is no mathematical contradiction here because we have three separate graphs and their solutions. We can, and do, talk about violations, not contradictions.

Note that in **SDL** we can derive Of and $O\neg f$ from (1), (2), (3) and we do have a problem, a contradiction, because we are working in a single same system.

Still, even for the equational approach, there is an intuitive difficulty here. The original $O\neg f$ is contradicted by $d \to Of$. The "contradiction" is that we offer a remedy for the violation d namley Of by violating $O\neg f$.

You might ask, why offer the remedy Of? Why not say keep the dog chained? Oc? The Oc remedy does not violate $O\neg f$.

The explanation that by having a dog (violating $O\neg d$) we created a new situation is rather weak, because having a fence is totally independent from having a dog, so we would expect that the remedy for having a dog will not affect $O\neg f$!

The important point is that the equational approach can identify such "inconsistencies" and can add constraints to avoid them if we so wish.

Remark 18. Let us adopt the view that once a violation is done by a fact then any type of new rules can be given. This settles the problems raised in Example 10. However, we have other problems. We still have to figure out a technical problem, namlely how to deal with several facts together. In the case of the Chisholm set there were no loops and so there was the natural obligation progression. We turned the set of facts into a sequence and separated the original graph (for the set of CTD wihout the facts) into a sequence of graphs, and this was our way of modelling the facts. When we have loops there is a problem of definition, how do we decompose the original graph when we have more than one fact? The next example will illustrate.

Example 11. This example has a loop and two facts. It will help us understand our modelling options in dealing with facts. Consider the following clauses. This is actually the Reykjavik paradox, see for example [13]:

1. There should be no dog
   $O\neg d$
2. There should be no fence
   $O\neg f$
3. If there is a dog then there should be a fence
   $d \to Of$
4. If there is a fence then there should be a dog
   $f \to Od$.

The figure for these clauses is Figure 26.

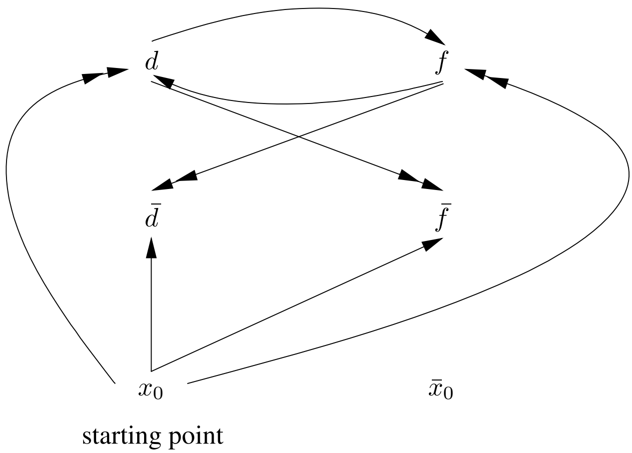


Fig. 26.

The soft equations solve this figure into $x_0 = \bar{d} = \bar{f} = 1$. $f = d = 0$. We now add the input that there is a dog and a fence.

5. d: dog, f: fence

The question is how to split Figure 26 in view of this input.

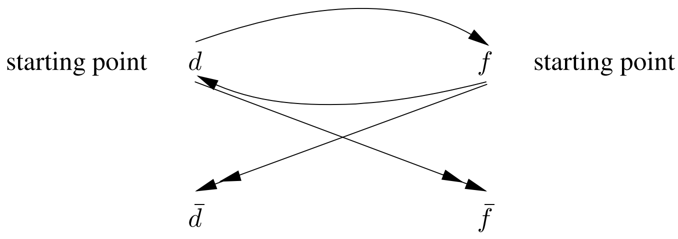


Fig. 27.

If we substitute $d = 1$ and $f = 1$ together and split, we get Figure 27, with two starting points.

Comparison with the original figure shows two violations of $O\neg d$ and $O\neg f$.

Let us now first add the fact d and then add the fact f.

When we add the fact d, Figure 26 split (actually is modified) into Figure 28. This figure happens to look just like Figure 27 with only d as a starting point. (Remember that any starting point x gets the equation $x = 1$.)

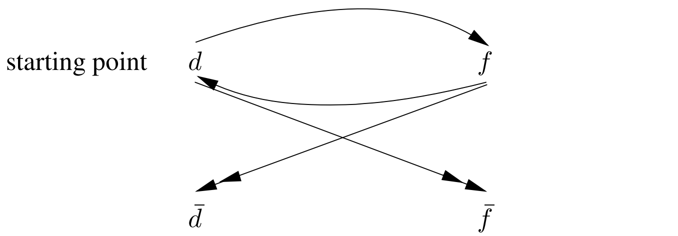


Fig. 28.

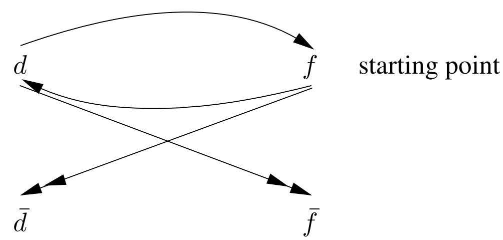


Fig. 29.

Adding now the additional fact f changes Figure 28 into Figure 29. In fact we would have got Figure 29 first, had we introduced the fact f first, and then added the fact d, we would have got Figure 29.

The difference between the sequencing is in how we perceive the violations.

The following is a summary.

Option 1. Introduce facts $\{d, f\}$ simultaneously. Get Figure 27, with two starting points. There are two violations, one of $O\neg f$ and one of $O\neg d$. This is recognised by comparing the solutions for the equations of Figure 26 with those of Figure 27.

Option 2. Introduce the fact d first. Figure 26 changes into Figure 28. Solving the equations for these two figures shows a violationof $O\neg d$ and a vioation of $O\neg f$, because f also gets $f = 1$ in the equations of Figure 28.

Option 2df. We now add to option 2d the further fact f. We get that Figure 28 becomes Figure 29. The solutions of the two figures are the soame, $f = d = 1$. So adding f gives no additional violation.

We thus see that adding $\{d, f\}$ together or first d and then f or (by symmetry) first f and then d all essentially agree and there is no problem. So where is the problem with simultaneous facts? See the next Example 12.

Example 12 (Example 11 continued). We continue the previous Example 11:

Let us try to add the facts $\{d, \neg f\}$ to the CTD set of Figure 26. Here we have a problem because we get Figure 30. In this figure both d and $\bar{f}$ are starting points. These two must solve to $d = \bar{f} = 1$. This is impossible in the way we set up the system. This means that it is inconsistent from the point of view of our semantics to add the facts $\{d, \neg f\}$ simultaneously in the semantics, or technically to have two starting points!

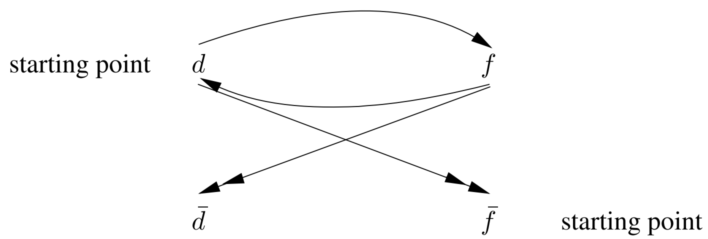


Fig. 30.

But we know that it is consisent and possible in reality to have a dog and no fence. So where did we go wrong in our semantic modelling? Mathematically the problem arises with making two nodes starting points. This means that we are making two variables equal to 1 at the same time. The equations cannot adjust and have a solution.[3]

The obvious remedy is to add the facts one at a time. Option 3d first adds d and then takes option 3d $\neg f$ and add $\neg f$ and in parallel, option 4$\neg f$ first adds the fact $\neg f$ and then take option 4$\neg fd$ and add the fact d. Let us see what we get doing these options and whether we can make sense of it.

[3] Remember when we substitute a fact we split the graph into two and so the equations change. We are not just substituting values into equations (in which case the order simultaneous or not does not matter), we are also changing the equations.

Recall what you do in Physics: If we have, for example, the equation $y = \sin x$ and we substitute for x a very small positive value, then we change the equation to $y = x$.

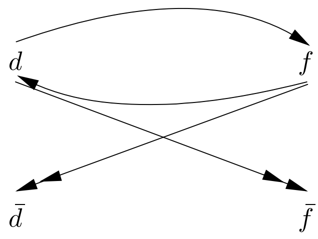


Fig. 31.

*Option 3*$d\neg f$. Adding the fact d would give us Figure 28 from Figure 26. We now add fact $\neg f$. This gives us Figure 31 from Figure 28.

Figure 31 violates Figure 28.

Option 3$\neg fd$. If we add the fact $\neg d$ first, we get Figure 31 from the original Figure 26.

If we now add the fact d, we get Figure 28.

The solution to the equations of this figure is $d = 1, f = 1, \bar{f} = 0$, but we already have the fact $\neg d$, so the $f = 1$ part cannot be accepted.

Summing up:

- facts $\{d, \neg f\}$ cannot be modelled simultaneously.
- First d then $\neg f$, we get that $\neg f$ violates $d \to Of$.
- first $\neg f$ then d, we get that $d \to Of$ cannot be implemented.

So the differences in sequencing the facts manifests itself as differences in taking a point of view of the sequencing of the violations.

The two views, when we have as additional data both d and $\neg f$, are therefore the following:

we view $d \to Of$ as taking precedent and $\neg f$ is violating it

or

we view $O\neg f$ as as taking precedence over $d \to Of$ and hence $d \to Of$ cannot be implemented.

3 Equational Semantics for General CTD Sets

We now give general definitions for general equational semantics for general CTD sets.

Definition 13.

1. *Let Q be a set of distinct atoms. Let $\bar{Q}$ be $\{\bar{a}|a \in Q\}$. Let $Q^* = Q \cup \bar{Q} \cup \{\top, \bot\}$. For $x \in \bar{Q}$, let $\bar{x}$ be x (i.e. $\bar{\bar{x}} = x$). Let $\bar{\top} = \bot$ and $\bar{\bot} = \top$.*

2. *A general CTD clause has the form* $x \to Oy$, *where* $x, y \in Q^*$ *and* $x \neq \bot, y \neq \bot, \top$.
3. *Given a set* $\mathbb{C}$ *of general CTD clauses let* $Q^*(\mathbb{C})$ *be the set* $\{x, \bar{x} | x$ *appears in a clause of* $\mathbb{C}\}$.
4. *Define two relations on* $Q^*(\mathbb{C}), \to$ *and* $\twoheadrightarrow$ *as follows:*
   - $x \to y$ *if the clause* $x \to Oy$ *is in* $\mathbb{C}$
   - $x \twoheadrightarrow y$ *if the clause* $x \to O\bar{y}$ *is in* $\mathbb{C}$
5. *Call the system* $\mathbb{G}(\mathbb{C}) = (Q^*(\mathbb{C}), \to, \twoheadrightarrow)$ *the graph of* $\mathbb{C}$.
6. *Let* $x \in Q^*(\mathbb{C})$. *Let*

$$E(x \to) = \{y | x \to y\}$$
$$E(\twoheadrightarrow x) = \{y | y \twoheadrightarrow x\}$$

Definition 14.

1. *Let* $\mathbb{C}$ *be a CTD set and let* $\mathbb{G}(\mathbb{C})$ *be its graph. Let* x *be a node in the graph. Let* $\mathbf{f}$ *be a function from* $Q^*(\mathbb{C})$ *into [0,1].*
   Define
   $$E^+(\twoheadrightarrow x, \mathbf{f}) = \{y | y \twoheadrightarrow x \textit{ and } \mathbf{f}(x) > 0\}.$$
2. *Let* $\mathbf{f}, x$ *be as in (1). We say* $\mathbf{f}$ *is a model of* $\mathbb{C}$ *if the following holds*
   (a) $\mathbf{f}(\top) = 1, \mathbf{f}(\bot) = 0$
   (b) $\mathbf{f}(\bar{x}) = 1 - \mathbf{f}(x)$
   (c) $\mathbf{f}(x) = \min(\{\mathbf{f}(y) | x \to y\} \cup \{1 - \mathbf{f}(z) | z \in E^+(\twoheadrightarrow x, \mathbf{f})\})$
3. *We say* $\mathbf{f}$ *is a {0,1} model of* $\mathbb{C}$ *if* $\mathbf{f}$ *is a model of* $\mathbb{C}$ *and*$\mathbf{f}$ *gives values in {0,1}.*

Example 13. Consider the set

1. $\top \to a$
2. $a \to \bar{a}$

This set has no models. However (2) alone has a model $\mathbf{f}(a) = 0, \mathbf{f}(\bar{a}) = 1$. The equations for (2) are: $a = 1 - \bar{a}, a = \min(1 - a, \bar{a})$.

The graph for (1) and (2) is Figure 32

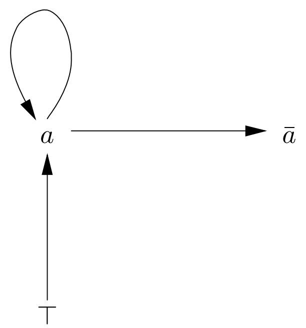


Fig. 32.

The graph for (2) alone is Figure 32 without the node $\top$.

Definition 15. *Let $\mathbb{C}$ be a CTD set. Let $\mathbb{G}(\mathbb{C}) = (Q^*, \to, \twoheadrightarrow)$ be its graph. Let $x \in Q^*$. We define the truncation graph $\mathbb{G}(\mathbb{C})_x$, as follows.*

1. *Let R^* be the reflexive and transitive closure of R where*

$$xRy = \mathrm{def}(\mathrm{x} \to \mathrm{y}) \vee (\mathrm{x} \twoheadrightarrow \mathrm{y}).$$

2. *Let Q^*_x be the set*

$$\{z | xR^*z \vee \bar{x}R^*z\} \cup \{\top, \bot\}$$

Let $\to_x = (\to \restriction Q_x) \cup \{\top \to x\}$.
Then

$$\mathbb{G}(\mathbb{C})_x = (Q^*_x, \to_x, \twoheadrightarrow \restriction Q^*_x).$$

3. *In words: the truncation of the graph at x is obtained by taking the part of the graph of all points reachable from x or $\bar{x}$ together with $\top$ and $\bot$ and adding $\top \to x$ to the graph.*

Example 14. Consider a m level Chisholm set as in Figure 11. The truncation of this figure at point $\bar{x}_3$ is essentially identicl with Figure 13. It is Figure 33. The difference is that we write "$\top \to \bar{x}_3$" instead of "$\bar{x}_3$ starting point". These two have the same effect on the equations namely that $\bar{x}_3 = 1$.

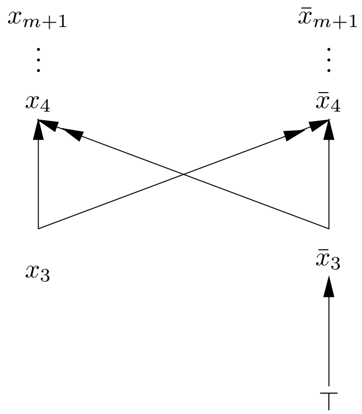


Fig. 33.

Definition 16. *Let $\mathbb{C}$ be a CTD set. Let $\mathbb{F}$ be a set of facts. We offer equational semantics for $(\mathbb{C}, \mathbb{F})$.*

1. *Let Ω be any ordering of $\mathbb{F}$.*

$$\Omega = (f_1, f_2, \ldots, f_k).$$

2. *Let* $\mathbb{G}(\mathbb{C})$ *be the graph of* $\mathbb{C}$ *and consider the following sequence of graphs and their respective equations.*

$$(\mathbb{G}(\mathbb{C}), \mathbb{G}(\mathbb{C})_{f_1}, \mathbb{G}(\mathbb{C})_{f_1 f_2}, \ldots, \mathbb{G}(\mathbb{C})_{f_1,\ldots,f_k})$$

We call the above sequence

$$\mathbb{G}(\mathbb{C})_{\Omega}.$$

We consider Ω *as a "point of view" of how to view the violation sequnce arising from the facts.*

3. *The full semantics for* $(\mathbb{C}, \mathbb{F})$ *is the family of all sequences* $\{\mathbb{G}(\mathbb{C})_{\Omega}\}$ *for all* Ω *orderings of* $\mathbb{F}$.

Remark 19. The CTD sets considered so far had the form $\pm x \to O \pm y$, where x and y are atomic. This remark expands our language allowing for x, y to be arbitrary propositional formulas. Our technical machinery of graphs and equations works just the same for this case. We can write in the graph $A, \bar{A}$ and then write the appropriate equations, and we use $\mathbf{h}_\mu$. For example the equation $\bar{A} = 1 - A$ becomes $\mathbf{h}_\mu(\bar{A}) = 1 - \mathbf{h}_\mu(A)$. Starting points A must satisfy $\mathbf{h}_\mu(A) = 1$, all the same as before. The only difference is that since the equations become implicit on the atoms, we may not have a solution.

In practice the way we approach such a CTD set is as follows: Let $\mathbb{C}_1$ be a set of CTD obligations of the form $\{A_i \to OB_i\}$. We pretend that A_i, B_i are all atomic. We do this by adding a new atomic constant $y(A)$, associated with every wff A. The set $\mathbb{C}_1 = \{A_i \to OB_i\}$ becomes the companion set $\mathbb{C}_2\{y(A_i) \to Oy(B_i)\}$. We now apply the graphs and equational approach to $\mathbb{C}_2$ and get a set of equations to be solved.

We add to this set of equations the further constraint equations

$$y(A_i) = \mathbf{h}_\mu(A_i)$$
$$y(B_i) = \mathbf{h}_\mu(B_i).$$

We now solve for the atomic propositions of the language.

We need to clarify one point in this set-up. What do we mean by facts $\mathbb{F}$? We need to take $\mathbb{F}$ as a propositional theory, it being the conjunction of some of the A_i. If we are given a set $\mathbb{C}_1$ of contrary to duty clauses of the form $A \to OB$ and facts $\mathbb{F}_1$, we check whether $\mathbb{F}_1 \vDash A$ in classical logic, (or in any other logic we use as a base. Note that if A_i are all atomic then it does not matter which logic we use as a base the consequence between conjunctions of atoms is always the same). If yes, then to the companion set $\mathbb{C}_2$ we add the fact $y(A)$. We thus get the companion set of facts $\mathbb{F}_2$ and we can carry on. This approach is perfectly compatible with the previous system where A, B were already atomic. The theory $\mathbb{F}$ is the conjunction of all the $\pm$ atoms in $\mathbb{F}$.

There is a slight problem here. When the formulas involved were atomic, a set of facts was a set of atoms $\mathbb{F}$, obtained by choose one of each pair $\{+x, -x\}$. So $\mathbb{F}$ was consistent. When the formulas involved are not atomic, even if we choose one of each pair $\{A, \neg A\}$, we may end up with a set $\mathbb{F}$ being inconsistent. We can require that we choose only consistent sets of facts and leave this requirement as an additional constraint.

This remark is going to be important when we compare our approach to that of Makinson and Torre's input output logic approach.

Example 15. To illustrate what we said in Remark 19, let us consider Figure 17.

The equations are listed in Example 8. Let us assume that in the figure we replace d by $y(D)$, where $D = d \vee c$. We get the equations involving $y(D)$ instead of D and get the solution as in Example 8, to be

$$x_0 = y(D) = \bar{f} = 1$$
$$\bar{x}_0 = y(\bar{D}) = f = 0.$$

Now we have the additional equation

$$\begin{aligned} y(D) &= \mathbf{h}_\mu(D) \\ &= \mathbf{h}_\mu(c \vee d) \\ &= \max(c, d) \end{aligned}$$

So we get $\max(c, d) = 1$ and we do have the solution with $d = 1, c = 1, ord = 0, c = 1, ord = 1, c = 0$.

Remark 20 (Miner Paradox [21,22]). We begin with a quote from Malte Willer in [22]

> Every adequate semantics for conditionals and deontic ought must offer a solution to the miners paradox about conditional obligations..... Here is the miners paradox. Ten miners are trapped either in shaft A or in shaft B, but we do not know which one. Water threatens to flood the shafts. We only have enough sand bags to block one shaft but not both. If one shaft is blocked, all of the water will go into the other shaft, killing every miner inside. If we block neither shaft, both will be partially flooded, killing one miner. [See Figure 34

Action	if miners in A	if miners in B
Block A	All saved	All drowned
Block B	All drowned	All saved
Block neither shaft	One drowned	One drowned

Fig. 34.

> Lacking any information about the miners exact whereabouts, it seems to say that
> 1. We ought to block neither shaft.
>
> However, we also accept that
> 2. If the miners are in shaft A, we ought to block shaft A,
> 3. If the miners are in shaft B, we ought to block shaft B.

But we also know that

4. Either the miners are in shaft A or they are in shaft B.

And (2)-(4) seem to entail

5. Either we ought to block shaft A or we ought to block shaft B, which contradicts (1).

Thus we have a paradox.

We formulate the Miners paradox as follows:

1. $\top \to O\neg\text{Block}A$
   $\top \to O\neg\text{Block}B$
2. Miners in $A \to O\text{Block}A$
3. Miners in $B \to O\text{Block}B$
4. Facts: Miners in $A \vee$ Miners in B.

The graph for (1)–(3) is Figure 35

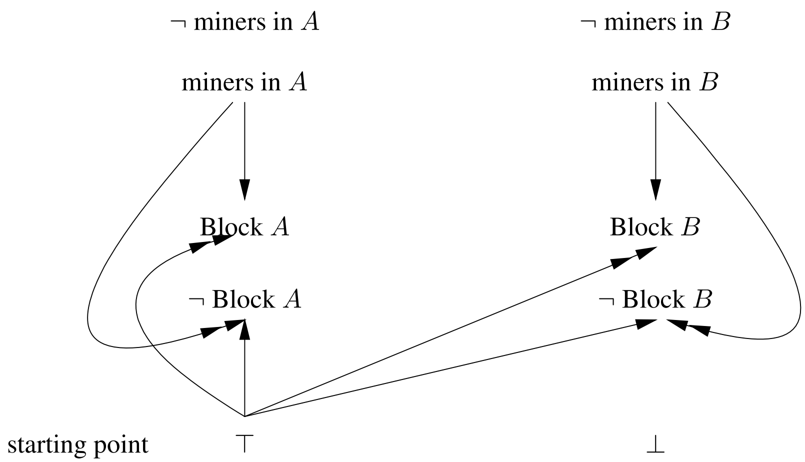


Fig. 35.

The Miners paradox arises because we want to detach using (2), (3) and (4) and get (5).

5. $O\text{Block}A \vee O\text{Block}B$

which contradicts (1).

However, according to our discussion, facts simply choose new starting points in the figure. The fact (4) is read as two possible sets of facts. Either the set of the fact that miner in A or the other possibility, the set containing miner in B. We thus get two possible graphs, Figure 36 and Figure 37.

We can see that there is no paradox here.

We conclude with a remark that we can solve the paradox directly using H. Reichenbach [24] reference points, without going through the general theory of this paper. See [23].

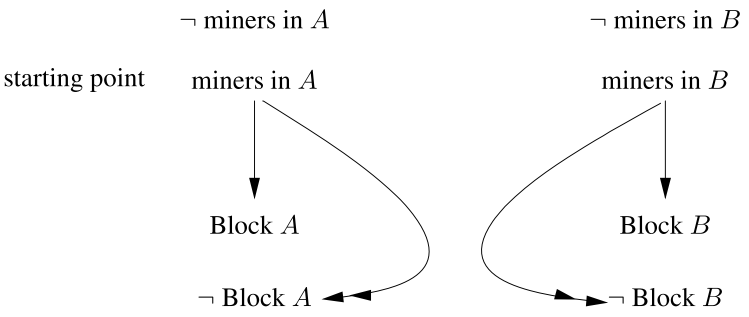


Fig. 36.

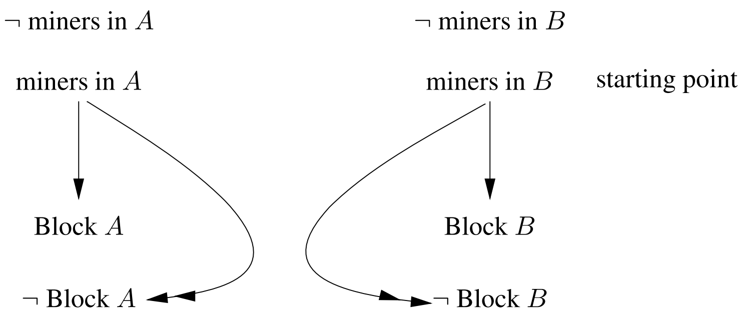


Fig. 37.

4 Proof Theory for CTDs

Our analysis in the previous sections suggest proof theory for sets of contrary to duty obligations. We use Gabbay's framework of labelled deductive systems [25].

We first explain intuitively our approach before giving formal definitions. Our starting point is Definition 13. The contrary to duty obligations according to this definition have the form $x \to Oy$, where x, y are atoms q or their negation $\neg q$ and x may be $\top$ and y is neither $\top$ nor $\bot$.

For our purpose we use the notation $x \Rightarrow y$. We also use labels annotating the obligations, and we write

$$t : x \Rightarrow y.$$

The label we use is the formula itself

$$t = (x \Rightarrow y).$$

Thus our CTD data for the purpose of proof theory has the form

$$(x \Rightarrow y) : x \Rightarrow y.$$

Given two CTD data items of the form

$$t : x \Rightarrow y; s : y \Rightarrow z$$

we can derive a new item

$$t * s : x \Rightarrow z$$

where $*$ is a concatenation of sequences. (Note that the end letter of t is the same as the beginning letter of s, so we can chain them.)

So we have the rule

$$\frac{(x \Rightarrow y) : x \Rightarrow y; (y \Rightarrow z) : y \Rightarrow z}{(x \Rightarrow y, y \Rightarrow z) : x \Rightarrow z}$$

It may be that we also have $(x \Rightarrow z) : x \Rightarrow z$ (i.e., the CTD set contains $x \to Oy, y \to Oz$ and $x \to Oz$), in which case $x \Rightarrow z$ will have two different labels, namely

$$t_1 = (x \Rightarrow y, y \Rightarrow z)$$
$$t_2 = (x \Rightarrow z).$$

We thus need to say that the proof theory allows for lables which are sets of chained labels (we shall give exact definitions later). So the label for $x \Rightarrow z$ would be $\{t_1, t_2\}$. There may be more labels $t_3, t_4, \ldots$ for $x \Rightarrow z$ depending on the original CTD set.

Suppose that in the above considerations $x = \top$. This means that our CTD set described above has the form $\{Oy$ (being $\top \to Oy), y \to Oz$ and $Oz\}$. By using the chaining rule we just described (and not mentioning any labels) we also get

$$\top \Rightarrow z, \top \Rightarrow y.$$

We can thus intuitively detach with $\top$ and get that our CTD set proves $\{y, z\}$. Notations $\vdash \{y, z\}$.

Alternatively, even if x were arbitrary, not necessarily $\top$, we can detach with x and write $x \vdash \{y, z\}$.

Of course when we use labels we will write

$$t : x \vdash \{s_1 : y, s_2 : z\}$$

the labels s_1, s_2 will contain in them the information of how y, z were derived from x.

To be precise, if for example,

$$\mathbb{C} = \{(x \Rightarrow y) : x \Rightarrow y, (y \Rightarrow z) : y \Rightarrow z, (x \Rightarrow z) : x \Rightarrow z\}.$$

We get

$$(x) : x \vdash_{\mathbb{C}} \{(x, x \Rightarrow y, y \Rightarrow z) : z, (x, x \Rightarrow z) : z, (x, x \Rightarrow y) : y\}.$$

Definition 17. *1. Let Q be a set of atoms. Let $\neg$ be a negation and let $\Rightarrow$ be a CTD implication symbol.*
A clause has the form $x \Rightarrow y$, where x is either $\top$ or atom q or $\neg q$ and y is either atom a or $\neg a$.

2. A basic label is either $(\top)$ or (q) or $(\neg q)$. (q atomic) or a clause $(x \Rightarrow y)$.

3. *A chain label is a sequence of the following form*

$$(x_0 \Rightarrow x_1, x_1 \Rightarrow x_2, \ldots, x_n \Rightarrow x_{n+1})$$

where $x_i \Rightarrow x_{i+1}$ are clauses. x_0 is called the initial element of the sequence and x_{n+1} is the end element.
4. *A set label is a set of chain labels.*
5. *A labelled CTD dataset $\mathbb{C}$ is a set of elements of the form $(x \Rightarrow y) : (x \Rightarrow y)$ where $x \Rightarrow y$ is a clause and $(x \Rightarrow y)$ is a basic label.*
6. *A fact has the form $(x) : x$ where x is either atom q or $\neg q$ or $\top$.*
A fact set $\mathbb{F}$ is a set of facts.

Definition 18. *Let $\mathbb{C}$ be a CTD dataset. We define the notions of*

$$\mathbb{C} \vdash_n t : x \Rightarrow y$$

where $n \geq 0, t$ a basic or chain label. This we do by induction on n.

We note that we may have $\mathbb{C} \vdash_n t : x \Rightarrow y$ hold for several different ns and different ts all depending on $\mathbb{C}$.

Case $n = 0$
$\mathbb{C} \vdash_0 t : x \Rightarrow y$ *if* $t = (x \Rightarrow y)$ *and* $(x \Rightarrow y) : x \Rightarrow y \in \mathbb{C}$.

Case $n = 1$
$\mathbb{C} \vdash_1 t : x \Rightarrow y$ *if for some* $x \Rightarrow w$ *we have* $(x \Rightarrow w) : x \Rightarrow w$ *in* $\mathbb{C}$ *and* $(w \Rightarrow y) : (w \Rightarrow y)$ *in* $\mathbb{C}$ *and* $t = (x \Rightarrow w, w \Rightarrow y)$.

Note that the initial element of t is x and the end element is y.

Case $n = m + 1$
Assume that $\mathbb{C} \vdash_m t : x \Rightarrow y$ has been defined and that in such cases the end element of t is y and the initial element of t is x.

Let $\mathbb{C} \vdash_{m+1} t : x \Rightarrow y$ hold if for some $t' : x \Rightarrow w$ we have $\mathbb{C} \vdash_m t' : x \Rightarrow w$ (and therefore the end element of t' is w and the initial element of t is x) and $(w \Rightarrow y) : w \Rightarrow y \in \mathbb{C}$ and $t = t' * (w \Rightarrow y)$, where $*$ is concatenation of sequences.

Definition 19. *Let $\mathbb{C}$ be a dataset and let $(x) : x$ be a fact. We write $\mathbb{C} \vdash^x_{n+1} t : y$ if for some $t : x \Rightarrow y$ we have $\mathbb{C} \vdash_n t : x \Rightarrow y$.*

We may also use the clearer notation

$$\mathbb{C} \vdash_{n+1} (x, t) : y.$$

Example 16. Let $\mathbb{C}$ be the set

$$(x \Rightarrow y) : x \Rightarrow y$$
$$(y \Rightarrow z) : y \Rightarrow z$$
$$(z \Rightarrow y) : z \Rightarrow y$$

Then

$$\mathbb{C} \vdash_0 (x \Rightarrow y) : x \Rightarrow y$$
$$\mathbb{C} \vdash_2 (x \Rightarrow y, y \Rightarrow z, z \Rightarrow y) : x \Rightarrow y$$
$$\mathbb{C} \vdash_1^x (x \Rightarrow y) : y$$
$$\mathbb{C} \vdash_3^x (x \Rightarrow y, y \Rightarrow z, z \Rightarrow y) : y$$

or using the clearer notation

$$\mathbb{C} \vdash_1 (x, (x \Rightarrow y)) : y$$
$$\mathbb{C} \vdash_3 (x, (x \Rightarrow y, y \Rightarrow z, z \Rightarrow y)) : y.$$

Note also that y can be proved with different labels in different ways.

Definition 20. *Let $\mathbb{C}$ be a dataset and let $\mathbb{F}$ be a set of facts. We define the notion of $\mathbb{C}, \mathbb{F} \vdash_n (z, t) : x$ where x, z are atomic or negation of atomic and z also possibly $z = \top$, as follows:*

Case $n = 0$
$\mathbb{C}, \mathbb{F} \vdash_0 (z, t) : x$ *if* $(x) : x \in \mathbb{F}$ *and* $(z, t) = (x)$.

Case $n = m + 1$
$\mathbb{C}, \mathbb{F} \vdash_{m+1} (z, t) : x$ *if* $\mathbb{C} \vdash_n t : z \Rightarrow x$ *and* $z = \top$ *or* $(z) : z \in \mathbb{F}$.

Example 17. We continue Example 16. We have

$$\mathbb{C}, \{(z) : z\} \vdash_0 (z) : z$$
$$\mathbb{C} \vdash_2 (z, (z \Rightarrow y)) : y$$
$$\mathbb{C} \vdash_4 (x, (x \Rightarrow y, y \Rightarrow z, z \Rightarrow y)) : y$$

Example 18. To illustrate the meaning of the notion of $\mathbb{C}, \mathbb{F} \vdash t : x$ let us look at the CTD set of Figure 26 (this is the Reykajavic set) with $d =$ dog and $f =$ fence:

1. $O\neg d$
   written as $(\top \Rightarrow \neg d) : \top \Rightarrow \neg d$.
2. $O\neg f$
   written as $(\top \Rightarrow \neg f) : \top \Rightarrow \neg f$.
3. $d \to Of$
   written as $(d \Rightarrow f) : d \Rightarrow f$.
4. $f \to Od$ written as $(f \Rightarrow d) : f \Rightarrow d$.

The above defines $\mathbb{C}$. Let the facts $\mathbb{F}$ be $(d) : d$ and $(\neg f) : \neg f$. We can equally write the facts as

$$(\top \Rightarrow d) : \top \Rightarrow d)$$
$$(\top \Rightarrow \neg f) : \top \Rightarrow \neg f.$$

(a) CTD point of view
   Let us first look at the contrary to duty set and the facts intuitively from the deontic point of view. The set says that we are not allowed to have neither a dog d nor a

fence f. So good behaviour must "prove" from $\mathbb{C}$ the two conclusions $\{\neg d, \neg f\}$. This is indeed done by

$$\mathbb{C} \vdash_1 (\top, (\top \Rightarrow \neg d)) : \neg d$$
$$\mathbb{C} \vdash_1 (\top, (\top \Rightarrow \neg f) : \neg f$$

The facts are that we have a dog (in violation of $\mathbb{C}$) and not a fence

$$\mathbb{F} = \{(d) : d, (\neg f) : \neg f\}.$$

So we can prove

$$\mathbb{C}, \mathbb{F} \vdash_0 (d) : d$$
$$\mathbb{C}, \mathbb{F}) \vdash_0 (\neg f) : \neg f$$

but we also have

$$\mathbb{C}, \mathbb{F} \vdash_1 (d, (d \Rightarrow f)) : f.$$

We can see that we have violations, and the labels tell us what violates what. Let us take the facts as a sequence. First we have a dog and then not a fence. Let $\mathbb{F}_d = \{(d) : d\}$ and $\mathbb{F}_{\neg f} = \{(\neg f) : \neg f\}$.
then

$$\mathbb{C}, \mathbb{F}_d \vdash_0 (d) : d$$

$$\mathbb{C}, \mathbb{F}_d \vdash_1 (d, (d \Rightarrow f)) : f$$

which violates

$$\mathbb{C} \vdash_0 (\top, (\top \Rightarrow \neg f)) : \neg f$$

but $\mathbb{F} = \mathbb{F}_d \cup \mathbb{F}_{\neg f}$, and so $\mathbb{F}$ viewed in this sequece (first d then $\neg f$) gives us a choice of points of view. Is the addition $\neg f$ a violation of the CTD dog $\rightarrow O$ fence or is it in accordance with the original $O\neg f$?
The problem here is that the remedy for the violation of $O\neg d$ by the fact d is $d \rightarrow Of$, which is a violation of another CTD namely $O\neg f$. One can say the remedy wins or one can say this rememdy is wrong, stick to $O\neg f$.
The important point about the proof system $\mathbb{C}, \mathbb{F} \vdash_n t : A$ is that we can get exactly all the information we need regarding facts and violations.

(b) Modal point of view
To emphasise the mechanical uniterpreted nature of the proof system let us give it a modal logic interpretaiton. We regard the labels as possible worlds and regard $*$ as indicating accessibility. We read $\mathbb{C}, \mathbb{F} \vdash_n t : A$ as $t \vDash A$, in the model $\mathbf{m}$ defined by $\mathbb{C}, \mathbb{F}$ i.e. $\mathbf{m} = \mathbf{m}(\mathbb{C}, \mathbb{F})$). The model of (a) above is shown in Figure 38
What holds at node t in Figure 38 is the end element of the sequence t.
The facts give us no contradiction, because they are true at different worlds. At $(\top, (\top \Rightarrow d))$ we have dog and so at $(\top, (\top \Rightarrow d, d \Rightarrow f))$ we have a fence while at $(\top, (\top \Rightarrow \neg f))$, we have no fence.
Inconsistency can only arise if we have $(x) : z$ and $(x) : \neg z$ or $t : x \Rightarrow y$ and $t : x \Rightarrow \neg y$ but we cannot express that in our language.

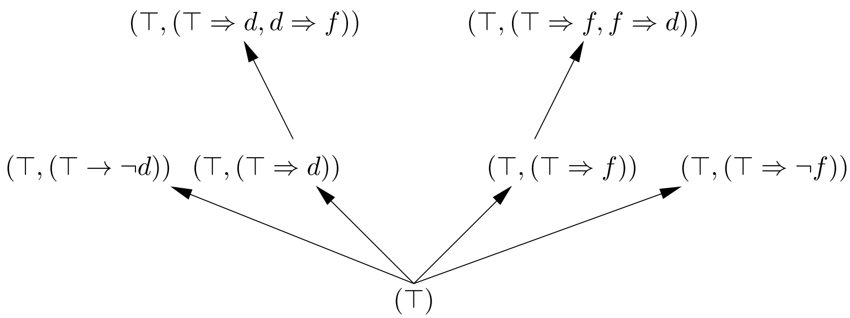


Fig. 38.

(c) The deductive view

This is a labelled deductive system view. We prove all we can from the system and to the extent that we get both A and $\neg A$ with different labels, we collect all labels and implement a *flattening policy*, to decide whether to adopt A or adopt $\neg A$.

Let us use, by way of example, the following flattening policy:

FP1 Longer labels win over shorter labels. (This means in CTD intrepretation that once an obligation is violated, the CTD has precedence.)

FP2 In case of same length labels, membership in $\mathbb{F}$ wins. This means we must accept the facts!

So according to this policy we have

$$(\top, (\top \Rightarrow d, d \Rightarrow f)) : f \text{ wins over } (\top, (\top \Rightarrow \neg f)) : \neg f$$

and

$$(\top, (\top \Rightarrow d)) : d \text{ wins over } (\top, (\top \Rightarrow \neg d)) : \neg d.$$

So we get the result $\{d, f\}$.

We can adopt the input-output policy of Makinson-Torre. We regard $\mathbb{C}$ as a set of pure mathematical input output pairs.

We examine each rule in $\mathbb{C}$ against the input $\mathbb{F}$. If it yields a contradictory output, we drop the rule.

The final result is obtained by closing the input under the remaining rules. So let us check:

$$\text{Input} : \{d, \neg f\}.$$

Rules in $\mathbb{C}$

$\top \Rightarrow \neg f$, OK

$\top \Rightarrow \neg d$, drop rule

$d \Rightarrow f$, drop rule

$f \Rightarrow d$, not applicable.

Result of closure: $\{d, \neg f\}$.

The input-output approach is neither proof theory not CTD. To see this add another rule

$$\neg d \Rightarrow b$$

where b is something completely different, consistent with $\pm d, \pm f$. This rule is not activated by the input $\{d, \neg f\}$. In the labelled approach we get b in our final set.

Problems of this kind have already been addressed by our approach of *compromise revision*, in 1999, see [25].

Example 19. Let us revisit the miners paradox of Remark 20 and use our proof theory.

We have the following data:

1. $\top \Rightarrow \neg$ Block A
2. $\top \Rightarrow \neg$ Block B
3. Miners in $A \Rightarrow$ Block A
4. Miners in $B \Rightarrow$ Block B
5. Fact: Miners in $A\vee$ miners in B

using ordinary logic.

We get from (2), (3) and (4)

5. Block $A\vee$ Block B

(5) contradicts (1).

Let us examine how we do this in our labelled system.

We have

1*. $(\top \Rightarrow$ Block $A) : \top \Rightarrow$ Bock A
$(\top \Rightarrow$ Block $B) : \top \Rightarrow$ Block B
2*. (miners in $A \Rightarrow$ Block A): miners in $A \Rightarrow$ Block A.
3*. (miners in $B \Rightarrow$ Block B): miners in $B \Rightarrow$ Block B
4*. (miners in $A\vee$ miners in B)': miners in $A\vee$ miners in B.

To do labelled proof theory we need to say how to chain the labels of disjunctions.

We do the obvious, we chain each disjunct. So if t is a label with end element $x \vee y$ and we have two rules $x \Rightarrow z$ and $y \Rightarrow w$ then we can chain

$$(t, (x \Rightarrow z, y \Rightarrow w))$$

So we have the following results using such chaining:
(1*), (2*), (3*), (4*) $\vdash_1 (\top, (\top \Rightarrow \neg$ Block $A)) : \neg$ Block A
(1*), (2*), (3*), (4*) $\vdash_1 (\top, (\top \Rightarrow \neg$ Block $B)) : \neg$ Block B
(1*), (2*), (3*), (4*) $\vdash_3$ ((miners$A \vee$ miners B), ((miners $A \Rightarrow$ Block A), (miners $B \Rightarrow$ Block B))) : Block $A \vee$ Block B

Clearly we have proofs of $\neg$ Block A, $\neg$ Block B and Block $A\vee$ Block B but with different labels! The labels represent levels of knowledge. We can use a flattening process on the labels, or we can leave it as is.

There is no paradox, because the conclusions are on different levels of knowledge.

This can be seen also if we write a classical logic like proof.

To prove Block $A \vee$ Block B from (2), (3), (4), we need to use subproofs. Whenever we use a subproof we regard the subproof as a higher level of knowledge. To see this consider the attempt to prove

$$\frac{(l1.1) \quad A \Rightarrow B}{(l1.2) \quad \neg B \Rightarrow \neg A}$$

We get $\neg B \Rightarrow \neg A$ from the subproof in Figure 39.

Let us now go back to the miners problem. The proof rules we have to use are

MP $\qquad \dfrac{A, A \Rightarrow B}{B}$

Outer Box

(11.2.1) Assume $\neg B$, show $\neg A$

To show $\neg A$ use subproof in Inner Box

Inner Box

(1 1.2.1.1) Assume A, show $\perp$

(11.2.1.2) We want to reiterate (11.1) $A \Rightarrow B$ and bring it here to do modus ponens and get B

(1 1.2.1.3) We want to reiterate (1 1.2.1) $\neg B$ and bring it here to get a contradiction.

To do these actions we need proof theoretic permissions and procedures, because moving assumptions across levels of knowledge, from outer box to inner box

Such procedures are part of the definition of the logic.

Fig. 39.

DE

$$\frac{\begin{array}{c} A \vee B \\ A \text{ proves } C \\ B \text{ proves } D \end{array}}{C \vee D}$$

RI We can reiterate positive (but not negative) wffs into subproofs.

The following is a proof using these rules:

Level 0

0.1a $\neg$ Block A. This is a negative assumption
0.1b $\neg$ Block B, negative assumption
0.2 miners in $A \Rightarrow$ Block A, assumption
0.3 miners in $B \Rightarrow$ Block B, assumption
0.4 miners in $A \vee$ miners in B
0.5 Block $A \vee$ Block B, would have followed from the proof in Figure 40, if there were no restriction rule RI. As it is the proof is blocked.

Box 1

1.1	Miners in $A \vee$ miners in B, reiteration of 0.4 into Box 1
1.2	Miners in $A \Rightarrow$ Block A, reiteration of 0.2 into Box 1.
1.3	Miners in $B \Rightarrow$ Block B, reiteration of 0.3 into Box 1
1.4	Block $A \vee$ Block B, from 1.1, 1.2, and 1.3 using DE
1.5	To get a contradiction we need to bring 0.1a $\neg$ Block A 0.1.b $\neg$ Block B as reiterations into Box 1. However, we cannot do so because these are negative information assumptions and cannot be reiterated

Fig. 40.

5 Comparing with Makinson and Torre's Input Output Logic

This section compares our work with Input Output logic , I/O, of Makinson and Torre. Our starting point is [19]. Let us introduce I/O using Makinson and Torre own words from [19].

> BEGIN QUOTE 1
> Input/output logic takes its origin in the study of conditional norms. These may express desired features of a situation, obligations under some legal, moral or practical code, goals, contingency plans, advice, etc. Typically they may be expressed in terms like: In such-and-such a situation, so-and-so should be the case, or ... should be brought about, or ... should be worked towards, or ... should be followed — these locutions corresponding roughly to the kinds of norm mentioned. To be more accurate, input/output logic has its source in a tension between the philosophy of norms and formal work of deontic logicians...
>
> Like every other approach to deontic logic, input/output logic must face the problem of accounting adequately for the behaviour of what are called 'contrary-to-duty' norms. The problem may be stated thus: given a set of norms to be applied, how should we determine which obligations are operative in a situation that already violates some among them. It appears that input/output logic provides a convenient platform for dealing with this problem by imposing consistency constraints on the generation of output.
>
> We do not treat conditional norms as bearing truth-values. They are not embedded in compound formulae using truth-functional connectives. To avoid all confusion, they are not even treated as formulae, but simply as ordered pairs (a,x) of purely boolean (or eventually first-order) formulae.
> Technically, a normative code is seen as a set G of conditional norms, i.e. a set of such ordered pairs (a, x). For each such pair, the body a is thought of as an input, representing some condition or situation, and the head x is thought of as an output, representing what the norm tells us to be desirable, obligatory or whatever in that situation. The task of logic is seen as a modest one. It is not to create or determine a distinguished set of norms, but rather to prepare information before it goes in as input to such a set G, to unpack output as it emerges and, if needed, coordinate the two in certain ways. A set G of conditional norms is thus seen as a transformation device, and the task of logic is to act as its 'secretarial assistant'.

Makinson and Torre adapt an example from Prakken and Sergot [4] to illustrate their use of input/output logic. We shall use the same example to compare their system with ours.

Example 20. We have the following two norms:

1. The cottage should not have a fence or a dog;
   $O\neg(f \vee d)$
   or equivalently

(a) $O\neg f$
(b) $O\neg d$

2. If it has a dog it must have both a fence and a warning sign.
$d \to O(f \wedge w)$
or equivalently
(c) $d \to Of$
(d) $d \to Ow$
In the notation of input/output logic the above data is written as
(e) $(\top, \neg(f \vee d))$
(f) $(d, f \wedge w)$.
Suppose further that we are in the situation that the cottage has a dog, in other words we have the fact:
3. Fact: d
thus violating the first norm.

The question we ask is: what are our current obligations? or in other words, how are we going to model this set? We know from our analysis in the previous section that a key to the problem is modelling the facts and that deontic logic gets into trouble because it does not have the means to pay attention to what we called the obligation progression.

Figures 41–43 describe our model, which is quite straight forward.

Let us see how Makinson and Torre handle this example.

The input output model will apply the data as input to the input output rules (f) and (e). This is the basic idea of Makinson and Torre for handling CTD obligations with facts.

Makinson and Torre realise that, and I quote again

BEGIN QUOTE 2
Unrestricted input/output logic gives
f: the cottage has a fence
and
w: the cottage has a warning sign.
Less convincingly, because unhelpful in the supposed situation, it also gives
$\neg d$: the cottage does not have a dog.
Even less convincingly, it gives
$\neg f$: the cottage does not have a fence,
which is the opposite of what we want. These results hold even for simple-minded output, ...

Makinson and Torre propose as a remedy to use constraints, namely to apply to the facts only those I/O rules which outputs are consistent with the facts. They say, and I quote again:

BEGIN QUOTE 3
Our strategy is to adapt a technique that is well known in the logic of belief change cut back the set of norms to just below the threshold of making the current situation contrary-to-duty. In effect, we carry out a contraction on the set G of given norms. Specifically, we look at the maximal subsets G' of G such

> that out(G', A) is consistent with input A. To illustrate this consider the cottage example, where $G = \{(t, \neg(f \vee d), (d, f \wedge w)\}$, with the contrary-to-duty input d. Using just simple minded output, G' has just one element $(d, f \wedge w)$ and so the output is just $f \wedge w$.

We note that this output corresponds to our Figure 43.

Makinson and Torre continue to say, a key paragraph showing the difference between our methods and theirs:

> BEGIN QUOTE 4
> Although the ... strategy is designed to deal with contrary-to-duty norms, its application turns out to be closely related to belief revision and nonmonotonic reasoning when the underlying input/output operation authorizes throughput More surprisingly, there are close connections with the default logic of Reiter, falling a little short of identity...

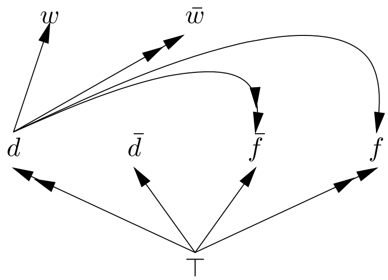


Fig. 41.

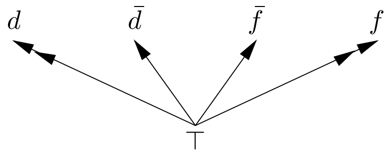


Fig. 42.

Let us, for the sake of comparison, consider the CTD sets of Figure 26 (this is actually the Reykajavik set of CTDs) and the facts as considered in Example 12. We have the following CTD (or equivalently the input output rules):

1. $(\top, \neg(d \vee f))$
2. (d, f)
3. (f, d)

The input is $A = d \wedge \neg f$.

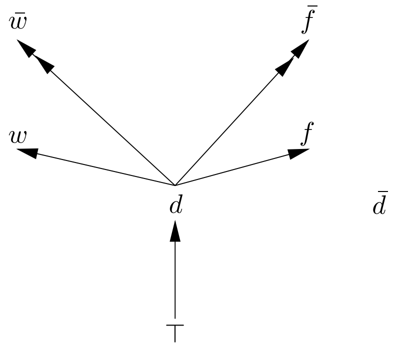


Fig. 43.

In this example the only rules (x, y) for which $A \vDash x$ are rules (1) and (2), but neither of their output is consistent with the input. So nothing can be done here. This corresponds to the lack of solution of our equations where we want to make both d and $\neg f$ the starting points. Our analysis in Example 12 however, gives a different result, because we first input d then $\neg f$ and in parallel, put $\neg f$ and then d.

So apart from the difference that input output logic is based on classical semantics for classical logic and we use equational semantics, there is also the difference that input output logic puts all the input in one go and detaches with all CTD rules whose output does not contradict it, while we use all possible sequencing of the input, inputting them one at a time. (To understand what 'one at a time' means, recall Remark 19 and Example 15.) There is here a significant difference in point of view. We take into account the obligation progression and given a set of facts as inputs, we match them against the obligation progression. In comparison, Input Output logic lumps all CTD as a set of input output engines and tries to plug the inputs into the engines in different ways and see what you get. The CTD clauses lose their Deontic identity and become just input output engines. See our analysis and comparison in part (c) of Example 18, where this point is clearly illustrated.

Let us do a further comparison. Consider the looping CTD set of Figure 4 which is analysed in Example 7. This has two input output rules

1. (a, b)
2. $(b, \neg a)$

Consider the two possible inputs

$$A = \neg a \wedge b$$

and

$$B = \neg a \wedge \neg b$$

A was a solution according to the soft approach option. B was a solution according to the mathematical approach option.

Using the input output approach, we can use $(b, \neg a)$ for A and we cannot use anything for B.

So there is compatibility here with the soft approach.

Let us summarise the comparison of our approach with the input output approach.

Com 1. We use equational sematics, I/O uses classical semantics.

Com 2. We rely on the obligation progression, breaking the input into sequence and modelling it using graphs. I/O does not do that, but uses the input all at once and taking maximal sets of CTDs (x, y) such that the input proves the xs and is consistent with the ys. The question whether it is possible to define violation progression from this is not clear. The I/O approach is a consequence relation/consistency approach. Our graph sequences and input facts sequences can also model action oriented/temporal (real time or imainary obligation progression 'time'). So for example we can model something like

$$f \to O\neg f$$

If you have a fence you should take it down.

Com 3. We remain faithful to the contrary to duty spirit, keeping our graphs and equations retain the CTD structure. I/O brought into their system significant AGM revision theory and turn I/O into a technical tool for revision theory and other nonmonotonic systems. See their quoted text 4.

Com 4. The connections are clear enough for us to say we can give equational semantics directly to input output logic, as it is, and never mind its connections with contrary to duty. Makinson and Torre defined input output logic, we have our equational approach, so we apply our approach to their logic directly. This is the subject of a separate paper.

6 Comparing with Governatori and Rotolo's Logic of Violations

We now compare with Governatori and Rotolo's paper [13]. This is an important paper which deserves more attention.

Governatori and Rotolo present a Gentzen system for reasoning with contrary-to-duty obligations. The intuition behind the system is that a contrary-to-duty is a special kind of normative exception. The logical machinery to formalise this idea is taken from substructural logics and it is based on the definition of a new non-classical connective capturing the notion of reparational obligation.

Given in our notation the following sequence of CTDs

$$\begin{aligned}
&A_1, \ldots, A_n \Rightarrow OB_1\\
&\neg B_1 \Rightarrow OB_2\\
&\neg B_1 \Rightarrow OB_3
\end{aligned}$$

They introduce a substructural connective and consider it as a sub-structural consequence relation without the structural rules of contraction, duplication, and exchange, and write the above sequence as

$$A_1, \ldots, A_n \Rightarrow B_1, \ldots, B_m.$$

The meaning is: the sequence $A_1, \ldots, A_n$ comports that B_1 is the case; but if B_1 is not satisfied, then B_2 should be the case; if both B_1 and B_2 are not the case, then B_3 should be satisfied, and so on. In a normative context, this means that the content of the obligation determined by the conditions $A_1, \ldots, A_n$. The As are the facts and they are not ordered. So in this respect Governatori and Rotolo approach are like the I/O approach.

Govenatori and Rotolo give proof theoretical rules for manipulating such sequents.

This approach is compatible with our approach in the sense that it relies on the obligation progression. It is also compatible with our proof theory of Section 4.

For the purpose of comparison, we need not go into details of their specific rules. it is enough to compare one or two cases.

Consider the CTD set represented by Figure 5. Since this set and figure is acyclic, Governatori and Rotolo can represent it by a theory containing several of their sequents . Each sequent will represent a maximal path in the figure. I don't think however that they can represent all possible paths. So the graph representation is a more powerful representation and we could and plan to present proof theory on graphs in a subsequent paper.

From my point of view , Governatori and Rotolo made a breakthrough in 2005 in the sense that they proposed to respect what I call the obligation (or violation) progression and their paper deserves more attention.

They use Gentzen type sequences which are written linearly, and are therefore restricted. We use plannar graphs (think of them as planar two dimensional Gentzen sequents) which are more powerful.

I am not sure how Governatori and Rotolo will deal with loops in general. They do find a way to deal with some loops for example I am sure they can handle the CTD of Figure 9,or of Figure 4, but I am not sure how they would deal with a general CTD set.

By the way, we used ordered sequences with hierarchical consequents in [18].

Governatori and Rotolo do not offer semantics for their system.

We offer equational semantics.

This means that we can offer equational semantics to their Gentzen system and indeed offer equational semantics to substructural logics in general.

This is a matter for another future paper.

Let us quote how they deal with the Chisholm paradox

> **Chisholm's Paradox.** The basic scenario depicted in Chisholm's paradox corresponds to the following implicit normative system:
>
> $$\{\vdash_O h, h \vdash_O i, \neg h \vdash_O \neg i\}$$
>
> plus the situation $s = \{\neg h\}$. First of all, note that the system does not determine in itself any normative contradiction. This can be checked by making explicit the normative system. In this perspective, a normative system consisting of the above norms can only allow for the following inference:
>
> $$\frac{\vdash_O h, \quad \neg h \vdash_O \neg i}{\vdash_O (h, \neg i)}$$

Thus, the explicit system is nothing but

$$\{h \vdash_O i, \vdash_O (h, \neg i)]\}.$$

It is easy to see that s is ideal (my words: i.e. no violations) wrt the first norm . On the other hand, while s is not ideal wrt $\vdash_O (h, \neg i)$, we do not know if it is sub-ideal (i.e. there are some violations but they are compensated by obeying the respective CTD) wrt such a norm. Then, we have to consider the two states of affairs $s_1 = \{\neg h, i\}$ and $s_2 = \{\neg h, \neg i\}$. It is immediate to see that s_1 is non-ideal (i.e. all violations throughout, no compensation) in the system, whereas s_2 is sub-ideal.

If so, given s, we can conclude that the normative system says that $\neg i$ ought to be the case.

7 Conclusion

We presented the equational approach for classical logic and presented graphs for General CTD sets which gave rise to equations . These equations provided semantics for general CTD sets.

The two aspects are independent of one another, though they are well matched.

We can take the graph representation and manipulate it using syntactical rules and this would proof theoretically model CTD's. Then we can give it semantics, either equational semantics or possible world semantics if we want.

We explained how we relate to Makinson and Torre's input output approach and Governatory and Rotolo's logic of violations approach.

The potential "output" from this comparison are the following possible future papers:

1. Equational semantics for input output logic
2. Equational semantics for substructural logics
3. Development of planar Gentzen systems (that would be a special case of labelled deductive systems)
4. Planar proof theory for input output logic (again, a special case of labelled deductive systems).
5. Proof theory and equational semantics for embedded CTD clauses of the form $x \to O(y \to Oz))$.

References

1. Gabbay, D.: An Equational Approach to Argumentation Networks. Argument and Computation 3(2-3), 87–142 (2012)
2. Gabbay, D.: Meta-Logical Investigations in Argumentation Networks. Research Monograph College publications, 770 p. (2013)
3. Gabbay, D.: Temporal deontic logic for the generalised Chisholm set of contrary to duty obligations. In: Ågotnes, T., Broersen, J., Elgesem, D. (eds.) DEON 2012. LNCS, vol. 7393, pp. 91–107. Springer, Heidelberg (2012)

4. Prakken, H., Sergot, M.J.: Contrary-to-duty obligations. Studia Logica 57(1), 91–115 (1996)
5. Carmo, J., Jones, A.J.I.: Deontic Logic and Contrary-to-duties. In: Gabbay, D.M., Guenthner, F. (eds.) Handbook of Philosophical Logic, vol. 8, pp. 265–343. Springer, Heidelberg (2002)
6. Prakken, H., Sergot, M.: Dyadic deontic logic and contrary to duty obligations. In: Nute, D. (ed.) Defeasible Deontic Logic, Synthese Library, pp. 223–262. Kluwer (1997)
7. de Boer, M., Gabbay, D., Parent, X., Slavkova, M.: Two dimensional deontic logic. Synthese 187(2), 623–660 (2012), doi:10.1007/s11229-010-9866-4
8. Jones, A.J.I., Pörn, I.: Ideality, sub-ideality and deontic logic. Synthese 65 (1985)
9. `http://en.wikipedia.org/wiki/T-norm`
10. Gabbay, D., Rodrigues, O.: Voting and Fuzzy Argumentation Networks. Submitted to Journal of Approximate Reasoning; Short version to appear in Proceedings of CLIMA 2012. Springer. New revised version now entitled: Equilibrium States on Numerical Argumentation Networks
11. Chisholm, R.M.: Contrary-to-duty imperatives and deontic logic. Analysis 24 (1963)
12. Gabbay, D., Gammaitoni, L., Sun, X.: The paradoxes of permission. An action based solution. To appear in Journal of Applied Logic
13. Governatori, G., Rotolo, A.: Logic of violations: a Gentzen system for reasoning with contrary-to-duty obligations. Australasian Journal of Logic 4, 193–215 (2005)
14. Governatori, G., Rotolo, A.: A Gentzen system for reasoning with contrary-to-duty obligations, a preliminary study. In: Jones, A.J.I., Horty, J. (eds.) Deon 2002, London, pp. 97–116 (May 2002)
15. Makinson, D., van der Torre, L.: Input/output logics. Journal of Philosophical Logic 29(4), 383–408 (2000)
16. Makinson, D., van der Torre, L.: Constraints for input/output logics. Journal of Philosophical Logic 30(2), 155–185 (2001)
17. Parent, X., Gabbay, D., van der Torre, L.: Intuitionistic Basis for Input/Output Logic. To appear springer volume in honour of David Makinson
18. Gabbay, D.M., Schlechta, K.: A Theory of Hierarchical Consequence and Conditionals. Journal of Logic, Language and Information 12(1), 3–32 (2010)
19. Makinson, D., van der Torre, L.: What is Input/Output Logic? In: ESSLLI 2001 (2001)
20. Gabbay, D., Strasser, C.: A dynamic proof theory for reasoning with conditional obligations on the basis of reactive graphs. Paper 477
21. Kolodny, N., MacFarlane, J.: Ifs and oughts. Journal of Philosophy 107(3), 115–143 (2010)
22. Willer, M.: A remark on iffy oughts. Journal of Philosophy 109(7), 449–461 (2012)
23. Baniasad, Z., Gabbay, D., Robaldo, L., van der Torre, L.: A solution to the miner paradox: A Reichenbach reference points approach (2013)
24. Reichenbach, H.: Elements of Symbolic Logic. Free Press, New York (1947)
25. Gabbay, D.M.: Labelled Deductive Systems. Oxford Logic Guides 33 (1996)
26. Gabbay, D.M.: Compromise update and revision: A position paper. In: Fronhoffer, B., Pareschi, R. (eds.) Dynamic Worlds. Applied Logic Series, vol. 12, pp. 111–148. Springer (1999)