Christoph Beierle
Carlo Meghini (Eds.)

LNCS 8367

Foundations of Information and Knowledge Systems

8th International Symposium, FoIKS 2014
Bordeaux, France, March 3-7, 2014
Proceedings

Springer

Christoph Beierle Carlo Meghini (Eds.)

Foundations of Information and Knowledge Systems

8th International Symposium, FoIKS 2014
Bordeaux, France, March 3-7, 2014
Proceedings

Springer

Volume Editors

Christoph Beierle
FernUniversität in Hagen
Fakultät für Mathematik und Informatik
Universitätstrasse 1
58084 Hagen, Germany
E-mail: christoph.beierle@fernuni-hagen.de

Carlo Meghini
Istituto di Scienza e Tecnologie della Informazione
Consiglio Nazionale delle Ricerche
Via G. Moruzzi 1
56124 Pisa, Italy
E-mail: carlo.meghini@isti.cnr.it

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-319-04938-0 e-ISBN 978-3-319-04939-7
DOI 10.1007/978-3-319-04939-7
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014930887

CR Subject Classification (1998): G.2, F.4.1-2, I.2.3-4, D.3

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Table of Contents

Invited Talks

Regular Articles

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.

- $h(A \wedge B) = 1$ *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 $\mathbf{n}$ 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 $\mathbf{n}$ 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 $\mathbf{n}$, and the function $\nu(x) = 1 - x$.

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

This is not the case for $\mathbf{m}$. 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 $\mathbf{n}$.

The case for $\mathbf{m}$ 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 $\mathbf{m}$, 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	$\neg A$	$A \wedge B$	$A \vee B$	$A \rightarrow B$
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*

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

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 $\rightarrow [0, 1]$ is an μ model of the theory if we have that* **f** *is a solution of the system of equations* $\mathbf{Eq}(\varDelta)$.

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

(b) *$\varDelta$ 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 $\{a = 1, b = 0\}$.

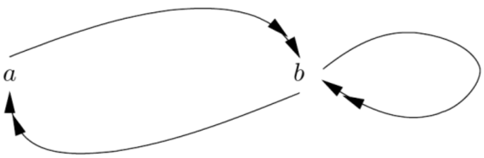
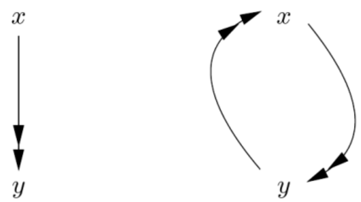


Fig. 2.

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).



(a) (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 $\mathbf{n}$. 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 $\mathbf{f}$ 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 $\to$ and the unary operator $\bigcirc$. We can write statements like

1. $\bigcirc\neg$ fence
   You should not have a fence
2. fence $\to \bigcirc$ 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

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

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 $S \neq \varnothing$ 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 $I(x) \neq \varnothing$ 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 **L**.
We expect the following to hold.

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

(b) *(2) and (3) are translated the same way, i.e., we translate the form:*
(23): $X \rightarrow$ 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* **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* **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) $p \rightarrow q$.
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) $O(p \rightarrow q)$

(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 $\neg y$".

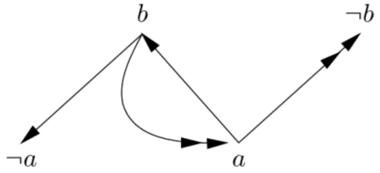


Fig. 4.

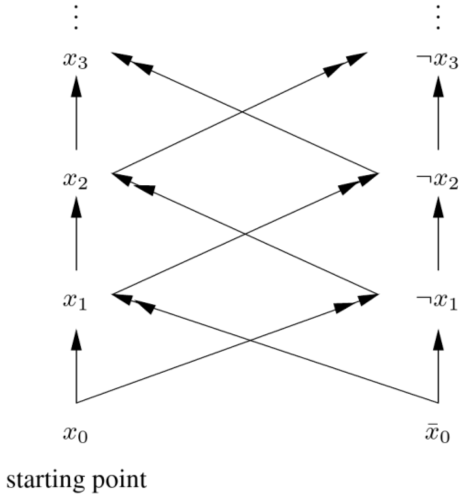


Fig. 5.

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 $\rightarrow$ 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 $(\pm x_0, \pm x_1, \pm x_2, \ldots)$.

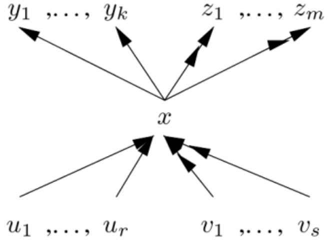


Fig. 6.

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

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

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

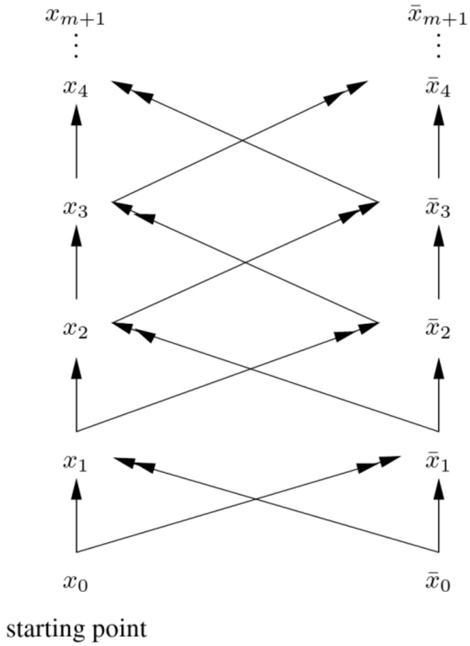


Fig. 7.

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

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

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*

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

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*

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

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 $1 \leq i \leq m$.

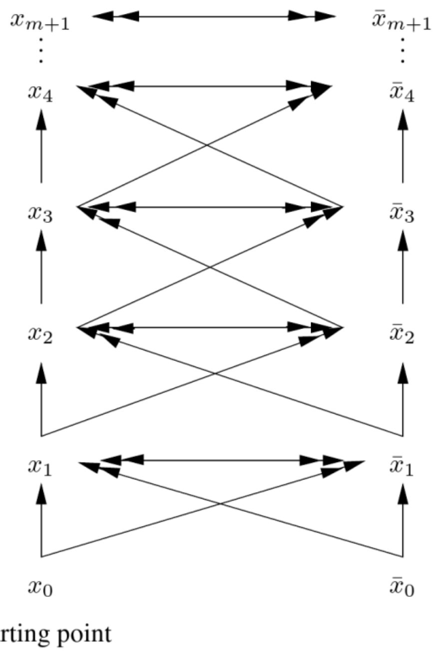


Fig. 11.

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

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

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 $i = 0, 1, \dots, m\}$. 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 $i = 0, 1, \dots, m\}$. 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 $(x_0, z_1, \dots, z_{m+1})$ 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

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

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 $(x_0, \bar{x}_1, x_2, \dots)$. 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.

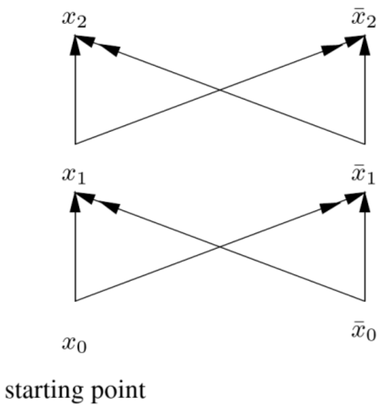


Fig. 12.

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*

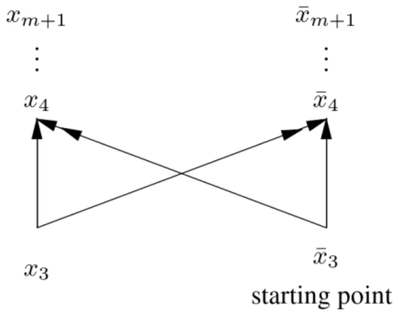


Fig. 13.

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).

image
not
available

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.

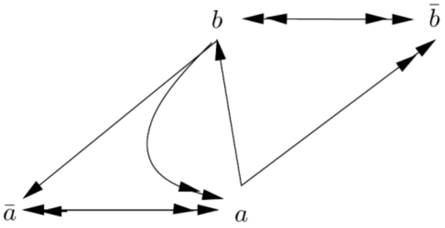


Fig. 16.

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

$$0 = \min(0, 0, 1)$$
$$1 = 1 - 0$$
$$0 = \min(1, 1 - 1)$$
$$1 = \min(1 - 0, 1 - 0).$$

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:

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

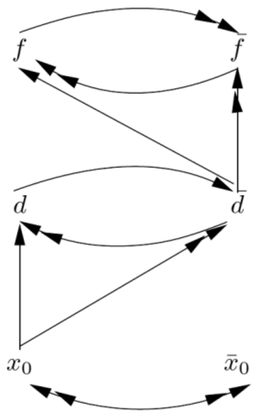


Fig. 17.

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

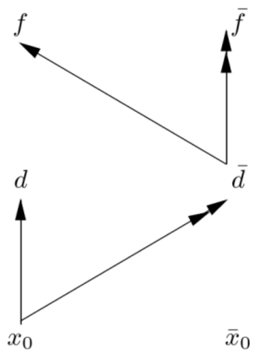


Fig. 18.

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.

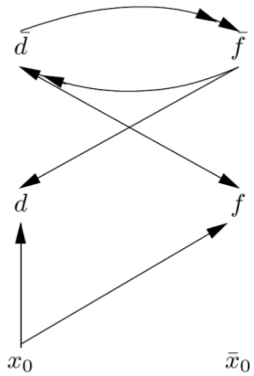


Fig. 19.

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

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

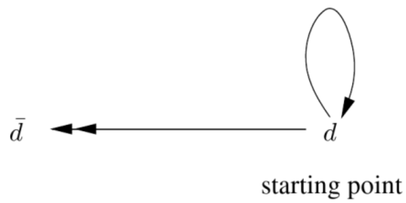


Fig. 22.

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
   $d \rightarrow Of$

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

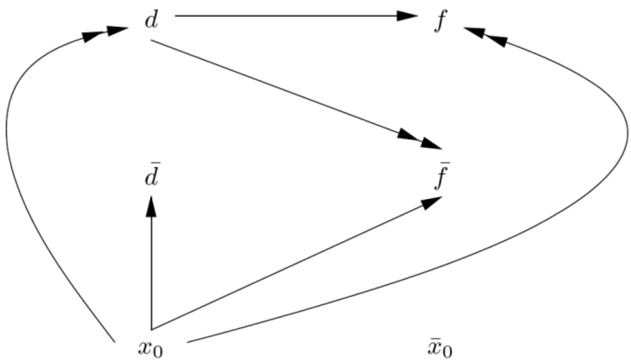


Fig. 23.

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

Author Index