



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

## ARCHIVIO ISTITUZIONALE DELLA RICERCA

### Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

Sequence semantics for normative agents

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

*Published Version:*

Sequence semantics for normative agents / Governatori, Guido; Olivieri, Francesco; Calardo, Erica; Rotolo, Antonino; Cristani, Matteo. - STAMPA. - 9862:(2016), pp. 230-246. [10.1007/978-3-319-44832-9\_14]

*Availability:*

This version is available at: <https://hdl.handle.net/11585/602594> since: 2019-03-14

*Published:*

DOI: [http://doi.org/10.1007/978-3-319-44832-9\\_14](http://doi.org/10.1007/978-3-319-44832-9_14)

*Terms of use:*

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).  
When citing, please refer to the published version.

(Article begins on next page)

This is a post-peer-review, pre-copyedit version of:

Governatori G., Olivieri F., Calardo E., Rotolo A., Cristani M. (2016) Sequence Semantics for Normative Agents. In: Baldoni M., Chopra A., Son T., Hirayama K., Torroni P. (eds) PRIMA 2016: Principles and Practice of Multi-Agent Systems. PRIMA 2016. Lecture Notes in Computer Science, vol 9862. Springer, Cham

The final authenticated version is available online at: [https://doi.org/10.1007/978-3-319-44832-9\\_14](https://doi.org/10.1007/978-3-319-44832-9_14)

This version is subjected to Springer Nature terms for reuse that can be found at: <https://www.springer.com/gb/open-access/authors-rights/aam-terms-v1>

# Sequence Semantics for Normative Agents

Guido Governatori<sup>1</sup>(✉), Francesco Olivieri<sup>2</sup>,  
Erica Calardo<sup>3</sup>, Antonino Rotolo<sup>3</sup>, and Matteo Cristani<sup>2</sup>

<sup>1</sup> Data61, CSIRO, Spring Hill, QLD, Australia

`guido.governatori@data61.csiro.au`

<sup>2</sup> University of Verona, Verona, Italy

<sup>3</sup> University of Bologna, Bologna, Italy

**Abstract.** We proposed a novel framework for the representation of goals and other mental-like attitudes in terms of degree of expected outcomes, where an outcome is an order of possible alternatives. The sequences of alternatives is modelled by a non-classical (substructural) operator. In this paper we provide a modal logic based axiomatisation of the intuition they propose, and we discuss some variants (in particular for the notion of social intention, intentions that are compliant with norms). Given that the outcome operator is substructural, we first propose a novel sequence semantics (a generalisation of possible world semantics) to model the outcome operator, and we prove that the axiomatisation is sound and complete with respect to the new semantics.

## 1 Introduction and Background

Normative Multi-Agent Systems (NorMAS) proved a powerful abstraction for the modelling of socio-technical systems [1]. A normative agent integrates two components: a rational agent component, often inspired by the well known BDI agent architecture, and the normative component to model the norms the agent is subject to.

The BDI architecture is a prominent approach to model rational agents. As is well-known, BDI agents are means-ends reasoners equipped with: (i) Desires, Goals, Intentions (or Tasks); (ii) a description of the current state of the environment (Beliefs); (iii) Actions. The key tenet of this architecture is that the agent's behaviour is the outcome of a rational balance among different mental states.

The framework we developed in [9] considers goals, desires, and intentions as facets of the same phenomenon (all of them being goal-like attitudes): the notion of *outcome*, which is simply something an agent would like or is expected to achieve. An advantage of the proposed framework is that it allows agents to compute different degrees of motivational attitudes, and degrees of commitment that take into account other factors, such as beliefs and norms.

While different schemas for generating and filtering agents’ outcomes are possible, the authors of [9] restricted themselves to schemas where they adopted the following principles:

- When an agent faces alternative outcomes in a given context, these outcomes are ranked in preference orderings;
- Mental attitudes are obtained from *outcomes*, which are ranked;
- Beliefs prevail over conflicting motivational attitudes, thus avoiding various cases of wishful thinking [3, 10];
- Norms and obligations are used to filter social motivational states (*social intentions*) and compliant agents [3, 6].

The motivational and deliberative components of agents are generated from preference orderings among outcomes. As done in other research areas (e.g., rational choice theory), we move with the idea that agents have preferences and choose which ones to attain the least of in given situations based on such preferences. Preferences involve outcomes and are explicitly represented in the syntax of the language for reasoning about agents, thus following the logical paradigm initially proposed in [2, 5] for the representation of preferences as explicit orders (sequences) of logical formulae and inference mechanisms to reason about them.

The combination of an agent’s mental attitudes with the factuality of the world defines her deliberative process, i.e., the objectives she decides to pursue. The agent may give up some of them to comply with the norms, if required. Indeed, many contexts may prevent the agent from achieving all of her objectives; the agent must then understand which objectives are mutually compatible with each other and choose which ones to attain the least of in given situations by ranking them in a preference ordering.

Consider, for instance, the following scenario. Alice is thinking what to do on Saturday afternoon. She has three alternatives: (i) she can visit John; (ii) she can visit her parents who live close to John’s place; or (iii) she can watch a movie at home. The alternative she likes the most is visiting John, while watching a movie is the least preferred. If John is not at home, there is no point for Alice to visit him. In this case, paying a visit to her parents becomes the “next best” option. Also, if visiting her parents is not possible, she settles for the last choice, that of staying home and watching a movie.

Suppose that Alice knows that John is actually away for the weekend. Since the most preferred option is no longer available, she decides to opt for the now best option, namely visiting her parents.

To represent the scenario above, we need to capture the preferences about her alternatives, and her beliefs about the world. To model preferences among several options, we build a sequence of alternatives that are preferred when the previous choices are no longer feasible. Normally, each set of alternatives is the result of a specific context determining under which conditions (premises) such a sequence of alternatives is considered.

Accordingly, we can represent Alice’s alternatives with the notation

$$\textit{Saturday} \rightarrow (\textit{visit\_John} \odot \textit{visit\_parents} \odot \textit{watch\_movie})$$

where the  $\odot$  operator is used to encode the preference of the agent over alternative outcomes. In this case *visit\_John* is the most preferred outcome, *visit\_parents* is the second best outcome (we can see it as the plan B of the agent), and *watch\_movie* is the least of the acceptable outcome for a Saturday afternoon. Notice that *do\_home\_chores* is not an “acceptable” outcome for Alice.

This intuition resembles the notion of contrary-to-duty obligations presented by [5], where a norm is represented by an expression like

$$drive\_car \rightarrow (\neg damage \otimes compensate \otimes foreclosure)$$

where the symbol “ $\otimes$ ” separates the alternatives. In this case, each element of the chain is the reparative obligation that shall come in force in case the immediate predecessor in the chain has been violated. Thus, the meaning of the formula above is that, if an agent drives a car, then she has the obligation not to cause any damage to others; if this happens, she is obliged to compensate; if she fails to compensate, there is an obligation of foreclosure.

In both examples, the sequences express a preference ordering among alternatives. Accordingly, *watch\_movie* and *foreclosure* are the last (and least) acceptable situations. Notice that while Alice’s alternatives come from a mental and a (inner) deliberation process, the use of  $\otimes$  describes situations of compliance with regards the environment the agent is situated in, and as such is up to the agent to comply with them, or not. Whilst both operators can be said to express preferences, who those preferences belong to differ: in the  $\odot$  case, it is the agent’s preferences, while in the  $\otimes$  case, it could be argued to be the stated preference of the institution/norms. These are not necessarily the same, as there are a number of instances in the real world of organisations breaking laws knowing that they will be fined: the institution prefers compliance over the fine, but the organisation may prefer paying the fine over complying with the law.

*Example 1.* Alice settled for visit her parents who live downtown, but the downtown part where they live is a traffic area restricted to residents and parking for not resident is forbidden on weekends. Alice would prefer to take her car to visit her parent to take public transports. The scenario can be represented by the formulas:

$$visit\_parents \rightarrow (parking\_downtown \odot public\_transport \odot pay\_fine) \quad (1)$$

$$weekend \rightarrow (\neg parking\_downtown \otimes pay\_fine) \quad (2)$$

In this case Alice has to balance her preference for driving her car to visit her parents and the prohibition to park downtown. But then she prefers to take a bus instead of paying a parking fine.

In the rest of the section, we shall illustrate the principles and intuitions relating sequences of alternatives (that is, outcome rules), beliefs, obligations, and how to use them to characterise different types of goal-like attitudes and degrees of commitment to outcomes: *desires*, *goals*, *intentions*, and *social intentions*.

*Desires as Acceptable Outcomes.* Desires are acceptable outcomes independently of whether they are compatible with other expected or acceptable outcomes. Let us contextualise the previous example to better explain the notion of desire by considering the following setting.

*Example 2.* Consider the following formulae:

$$Saturday \tag{3}$$

$$John\_sick \tag{4}$$

$$Saturday \rightarrow (visit\_John \odot visit\_parents \odot watch\_movie) \tag{5}$$

$$John\_sick \rightarrow (\neg visit\_John \odot short\_visit). \tag{6}$$

The agent has both *visit\_John* and its opposite as acceptable outcomes.

*Goals as Preferred Outcomes.* We consider a goal as the preferred desire in a chain.

In the situation described by Example 2, *visit\_parents* and *short\_visit* are the goals we can obtain: a desire is a goal only if it is compatible with other desires.

### 1.1 Two Degrees of Commitment: Intentions and Social Intentions

The next issue is to clarify which are the acceptable outcomes for an agent to commit to. Naturally, if the agent values some outcomes more than others, she should strive for the best, in other words, for the most preferred outcomes (goals).

Consider a formula  $b_1 \odot b_2 \odot b_3$ . Let us examine a first case where the agent should commit to the outcome she values the most, that is  $b_1$ . But what if the agent *believes* that  $b_1$  cannot be achieved in the environment where she is currently situated in, or she knows that  $\neg b_1$  holds? Committing to  $b_1$  would result in a waste of the agent's resources; rationally, she should target the next best outcome  $b_2$ . Accordingly, the agent derives  $b_2$  as her *intention*. *An intention is an acceptable outcome which does not conflict with the beliefs describing the environment.*

Suppose now that  $b_2$  is *forbidden*, and that the agent is social (a social agent is an agent not knowingly committing to anything that is forbidden [6]). Once again, the agent has to lower her expectation and settle for  $b_3$ , which is one of her *social intentions*. *A social intention is an intention which does not violate any norm.*

### 1.2 The Contribution of This Paper

The above notions have been formalised by [9] in the context of Defeasible Logic. That model was motivated by computational concerns: the proposed logic for agents' desires, goals, and intentions has in fact linear complexity. However, the approach was only proof-theoretic whereas no semantic model-theoretic analysis of the operators  $\otimes$  and  $\odot$  has been provided. In this paper, we work on the combination of these operators in the context of classical propositional logic and fill

the gap by defining a suitable possible-world semantics for them. Such semantics is a sequence-based non-normal one extending and generalising neighbourhood models for classical modal logics.

The layout of the paper is as follows: Sect. 2 presents the language of our logic; Sect. 3 discusses some axiom schemata governing the behaviour and the interactions of the various operators; Sect. 4 describes an account of the semantics for the logic while Sect. 5 offers some relevant completeness results. Some conclusions end the paper.

## 2 Language

The language consists of a countable set of atomic formulae. Well-formed formulae are then defined using the typical Boolean connectives, the  $n$ -ary connectives  $\otimes$  and  $\odot$ , and the modal (deontic) operators  $\mathbf{O}$  for obligation,  $\mathbf{B}$  for beliefs,  $\mathbf{D}$  for desires,  $\mathbf{G}$  for goals,  $\mathbf{I}$  for intentions, and  $\mathbf{SI}$  for social intentions. The intended reading of  $\otimes$  is that it encodes a sequence of obligations where, each obligation is meant to compensate the violation of the previous obligation. The intuition behind  $\odot$  is to model ordered lists of preferred outcomes.

Let  $\mathcal{L}$  be a language consisting of a countable set of propositional letters  $Prop = \{p_1, p_2, \dots\}$ , the propositional constant  $\perp$ , round brackets, the boolean connective  $\rightarrow$ , the unary operators  $\mathbf{O}$ ,  $\mathbf{B}$ ,  $\mathbf{D}$ ,  $\mathbf{G}$ ,  $\mathbf{I}$ , and  $\mathbf{SI}$ , the set of  $n$ -ary operators  $\otimes^n$  for  $n \in \mathbb{N}^+$  and the set of  $n$ -ary operators  $\odot^n$  for  $n \in \mathbb{N}^+$ .

**Definition 1 (Well Formed Formulae).** *Well formed formulae (wffs) are defined as follows:*

- Any propositional letter  $p \in Prop$  and  $\perp$  are wffs;
- If  $a$  and  $b$  are wffs, then  $a \rightarrow b$  is a wff;
- If  $a$  is a wff and no operator  $\otimes^m$ ,  $\odot^m$ ,  $\mathbf{O}$ ,  $\mathbf{B}$ ,  $\mathbf{D}$ ,  $\mathbf{G}$ ,  $\mathbf{I}$ , and  $\mathbf{SI}$  occurs in  $a$ , then  $\mathbf{O}a$ ,  $\mathbf{B}a$ ,  $\mathbf{D}a$ ,  $\mathbf{G}a$ ,  $\mathbf{I}a$ , and  $\mathbf{SI}a$  are a wff;
- If  $a_1, \dots, a_n$  are wffs and no operator  $\otimes^m$ ,  $\odot^m$ ,  $\mathbf{O}$ ,  $\mathbf{B}$ ,  $\mathbf{D}$ ,  $\mathbf{G}$ ,  $\mathbf{I}$ , and  $\mathbf{SI}$  occurs in any of them, then  $a_1 \otimes^n \dots \otimes^n a_n$  and  $a_1 \odot^n \dots \odot^n a_n$  are wffs, where  $n \in \mathbb{N}^+$ ;<sup>1</sup>
- Nothing else is a wff.

We use  $WFF$  to denote the set of well formed formulae.

Other Boolean operators are defined in the standard way, in particular  $\neg a =_{def} a \rightarrow \perp$  and  $\top =_{def} \perp \rightarrow \perp$ .

We say that any formula  $a_1 \otimes \dots \otimes a_n$  is an  $\otimes$ -chain, while  $a_1 \odot \dots \odot a_n$  is an  $\odot$ -chain; also the negation of an  $\otimes$ -chain (resp.  $\odot$ -chain) is an  $\otimes$ -chain (resp.  $\odot$ -chain). The formation rules allow us to have  $\otimes$ -chain and  $\odot$ -chain of any (finite) length, and the arity of the operator is equal to number of elements in the chain; we hence drop the index  $m$  from  $\otimes^m$  and  $\odot^m$ . Moreover, we use the prefix notation  $\bigotimes_{i=j}^n a_i$  for  $a_j \otimes \dots \otimes a_n$  and  $\bigodot_{i=j}^n a_i$  for  $a_j \odot \dots \odot a_n$ .

<sup>1</sup> We use the prefix forms  $\otimes^1 a$  and  $\oplus^1 a$  for the case of  $n = 1$ .

In addition, we use the following notation:  $\bigotimes_{i=j}^n a_i \otimes b \otimes \bigotimes_{k=l}^m c_k$ , where  $j, l \in \{0, 1\}$ . The “ $a$ ” part and “ $c$ ” part are optional, i.e., they are empty when  $j = 0$  or  $l = 0$ , respectively. Otherwise the expression stands for the following chain of  $n + 1 + m$  elements:  $a_1 \otimes \dots \otimes a_n \otimes b \otimes c_1 \otimes \dots \otimes c_m$ . The same reasoning holds for  $\bigodot_{i=j}^n a_i \odot b \odot \bigodot_{k=l}^m c_k$ .

### 3 Axiomatisation for Norms, Beliefs and Outcomes

The aim of this section is to discuss the intuitions behind some principles governing the behaviour and the interactions of the various operators. These principles are captured by axioms or inference rules.

#### 3.1 Basic Axioms and Inference Rules

In this paper, we assume classical propositional logic, CPC, as the underlying logic. The first principle is that of syntax independence or, in other terms, that the operators are closed under logical equivalence. To this end, all the logics have the following inference rules:

$$\frac{a \equiv b}{\Box a \equiv \Box b} \quad (\Box\text{-RE})$$

with  $\Box \in \{O, B, D, G, I, SI\}$ .

$$\frac{\bigwedge_{i=1}^n (a_i \equiv b_i)}{\bigotimes_{i=1}^n a_i \equiv \bigotimes_{i=1}^n b_i} \otimes\text{-RE} \quad \frac{\bigwedge_{i=1}^n (a_i \equiv b_i)}{\bigodot_{i=1}^n a_i \equiv \bigodot_{i=1}^n b_i} \odot\text{-RE} \quad (\otimes \text{ and } \odot\text{-RE})$$

Consider the  $\otimes$  chain  $a \otimes b \otimes a \otimes c$ . The meaning of the chain above is that  $a$  is obligatory, but if  $a$  is violated (meaning that  $\neg a$  holds) then  $b$  is obligatory. If also  $b$  is violated, then  $a$  becomes obligatory. But we already know that we will incur in the violation of it, since  $\neg a$  holds. We thus have the obligation of  $c$ . However, this is what we want to obtain from the  $\otimes$ -chain:  $a \otimes b \otimes c$ .

Now, consider the  $\odot$ -chain  $a \odot b \odot a \odot c$ . The intuitive reading is that  $a$  should be the most preferred outcome, while  $b$  is the second best in case  $a$  is not possible to achieve. However, if also  $b$  is not attainable, then  $a$  should be the agent’s third best choice. Nevertheless, we have already established that this is not possible, and we thus have  $c$  as the following preferred outcome.

The above example shows that duplications of formulas in  $\otimes$ -chains and  $\odot$ -chains do not contribute to the meaning of the chains themselves. This motivates us to adopt the following axioms to remove (resp., introduce) an element from (to) a chain if an equivalent formula occurs on the left of it.

$$\bigotimes_{i=1}^n a_i \equiv \bigotimes_{i=1}^{k-1} a_i \otimes \bigotimes_{i=k+1}^n a_i \text{ where } a_j \equiv a_k, j < k \quad (\otimes\text{-contraction})$$

$$\bigodot_{i=1}^n a_i \equiv \bigodot_{i=1}^{k-1} a_i \odot \bigodot_{i=k+1}^n a_i \text{ where } a_j \equiv a_k, j < k \quad (\odot\text{-contraction})$$



Given that we use classical propositional logic as the underlying logic, it is not possible that an  $\otimes$ -chain ( $\odot$ -chain) and its negation hold at the same time. What about when  $\otimes$ -chains like  $a \otimes b \otimes c$  and  $\neg(a \otimes b)$  hold. In this case, the first chain states that  $a$  is obligatory and its violation is compensated by  $b$ , which in turn is itself obligatory and it is compensated by  $c$ . The second expression states that ‘either it is not the case that  $a$  is obligatory, but if it is so, then its violation is not compensated by  $b$ ’. Accordingly, the combination of the two expressions should result in a contradiction (a similar argument can be made for  $\odot$ -chains). To ensure this, we must assume the following axioms that allow us to derive, given a chain, all its sub-chains with the same initial element(s).

$$\begin{aligned} a_1 \otimes \cdots \otimes a_n &\rightarrow a_1 \otimes \cdots \otimes a_{n-1}, \quad n \geq 2 && (\otimes\text{-shortening}) \\ a_1 \odot \cdots \odot a_n &\rightarrow a_1 \odot \cdots \odot a_{n-1}, \quad n \geq 2 && (\odot\text{-shortening}) \end{aligned}$$

If Alice prefers to visit John to visit her parents to watch a movie to spend her Saturday afternoon she prefer to visit John to visit her parents. Thus, we can derive

$$visit\_John \odot visit\_parents$$

from

$$visit\_John \odot visit\_parents \odot watch\_movie.$$

### 3.2 Axioms for Obligations and Mental Attitudes

In the previous section we proposed the basic axioms for a logic of norms and goals. In this section, we address the relationships between  $\otimes$ ,  $\odot$ , obligations, beliefs and mental attitudes;

In this paper we assume that the description of the environment in which an agent is situated is given by a set of propositional formulas. The agent is rational in the sense that the agent is able to reason with the formulas using classical propositional logic as the underlying logic. We further assume that the provided description of the environment is truthful and the agent knows it. Accordingly, we do not have to introduce a further modal operator to describe the knowledge of an agent. The second principle of rationality for the agents is that they do not form self-inconsistent beliefs and goal-like mental attitudes, and similarly for the norms (obligations) they are subject to. Namely, we assume that all unary modal operators are internally consistent. Internal consistency of beliefs, obligations, and goal-like mental attitudes is expressed by the following axiom:

$$\neg \Box \perp \tag{(\perp-\Box)}$$

with  $\Box = \{O, B, D, G, I, S\}$ .

Similarly, rational agents are expected to avoid conflictual beliefs and individual mental-attitudes (but desires), and there are no norms that make something obligatory and forbidden at the same time. This property is called external consistency and it is modelled by the following axiom:

$$\Box a \rightarrow \neg \Box \neg a \tag{(D)}$$

with  $\square = \{\mathbf{B}, \mathbf{O}, \mathbf{G}, \mathbf{I}, \mathbf{SI}\}$ . As we discussed in [9] we do not assume this external consistency for desires. Thus  $\mathbf{D}a$  and  $\mathbf{D}\neg a$  is consistent within our framework. For example, Alice may desire to visit Bob while, at the same time and for other reasons, she might not to. Visiting Bob and not visiting him are two possible outcomes for Alice deliberation, and she can use other information to determine what course of action she commits to. Hence, it is rational for her to derive that she has both desires (they are two acceptable and viable outcomes).

In addition the agent can have a set of beliefs of how the environment is (for example, about what is not given by the explicit description of the environment or can be inferred from it using classical propositional logic). Thus, to represent the agent belief we use the  $\mathbf{B}$  operator.  $\mathbf{B}$  is axiomatised as a normal  $\mathbf{KD45}$  operator, with the standard axioms for positive and negative introspections (Axioms 4 and 5). Accordingly the axioms for  $\mathbf{B}$  are:

$$\mathbf{B}a \rightarrow \mathbf{B}\mathbf{B}a \quad (4)$$

$$\neg \mathbf{B}\neg a \rightarrow \mathbf{B}\neg \mathbf{B}\neg a \quad (5)$$

$$\mathbf{B}(a \rightarrow b) \rightarrow (\mathbf{B}a \rightarrow \mathbf{B}b) \quad (\mathbf{K})$$

$$\mathbf{B}\top \quad (\mathbf{N})$$

For the other modal operators, we establish that they are regular, namely, that the following inference  $\square$ -RR holds for them

$$\frac{a_1 \wedge \dots \wedge a_n \rightarrow b}{\square a_1 \wedge \dots \wedge \square a_n \rightarrow \square b} \quad (\square\text{-RR})$$

for  $\square \in \{\mathbf{O}, \mathbf{D}, \mathbf{G}, \mathbf{I}, \mathbf{SI}\}$ . As explained in [6]  $\square$ -RR allows us to model, for example, the notion of intentionality: Suppose that an agent knows that  $a_1 \wedge \dots \wedge a_n \rightarrow b$  (or in other terms, that the implication is a property of the environment in which the agent is situated). This means, that in the given environment,  $b$  is an unavoidable consequence of  $a_1, \dots, a_n$ . Therefore, if the agents intend all the  $a_i$ s, i.e.,  $\mathbf{I}a_i$ , then the agent knows that in case she is successful in achieving all the  $a_i$ s, then she will bring about  $b$ , thus committing herself to the  $a_i$ s indirectly commits herself to  $b$  as well. Thus,  $\square$ -RR allows us to derive  $\mathbf{I}b$ .

We can now move to the analysis of the axioms relating norms ( $\otimes$ -chains) and obligations. In this paper we follow the analysis proposed in [8] for the relationships between norms and obligations. Thus the first axiom we consider is:

$$a_1 \otimes \dots \otimes a_n \rightarrow \mathbf{O}a_1 \quad (\otimes\text{-O})$$

For instance, given the “*driving\_car*” example above, we can obtain the first element as the current obligation ( $\mathbf{O}\neg$ *damage*). Furthermore, we say that if the first element does not hold, we can infer the obligation of the second element. For example

$$a_1 \otimes \cdots \otimes a_n \wedge \neg a_1 \rightarrow \mathcal{O}a_2. \quad (7)$$

In this occasion (always referring to the “*driving\_car*” example), we actually did cause some damage, thus the new obligation in force is to compensate for it (*Ocompensate*).

Moreover, we argued that we can repeat the same procedure. This leads us to generalise (7) for the axiom that expresses the detachment principle for  $\otimes$ -chains and factual statements about the opposites of the first  $k$  elements of an  $\otimes$ -chain.

$$a_1 \otimes \cdots \otimes a_n \wedge \bigwedge_{i=1}^{k < n} \neg a_i \rightarrow \mathcal{O}a_{k+1} \quad (\mathcal{O}\text{-detachment})$$

For alternative axiomatisation of the relationships between obligations and norms see [8].

As stated before, desires are expected or acceptable outcomes, independently of whether they are compatible with other expected or acceptable outcomes. As such, given an  $\odot$ -chain, we consider each element to be desired by the agent since she has expressed a preference order on such a chain and she thus considers all of them to be acceptable. This is expressed by the following axiom.

$$a_1 \odot \cdots \odot a_n \rightarrow \mathcal{D}a_i, \text{ with } 1 \leq i \leq n \quad (\odot\text{-D})$$

According to this axioms, given the preference of Alice her “desires”, i.e., the outcomes that she considers acceptable are *visit\_John*, *visit\_parents* and *watch\_movie*, but she has no desire to her household chores.

Given an  $\odot$ -chain, a goal represents what should be the most preferred outcome for the agent, provided that the agent does not have the opposite desire.

$$a_1 \odot \cdots \odot a_n \wedge \neg \mathcal{D}\neg a_1 \rightarrow \mathcal{G}a_1 \quad (\odot\text{-G}_1)$$

Again, if an agent has opposite desires then, given an  $\odot$ -chain, the goal is the first element such that the agent does not desire the opposite. By rationality, if for each element of such a chain, the agent has expressed an opposite desire, then any element of the chain cannot represent a goal.

$$a_1 \odot \cdots \odot a_n \wedge \bigwedge_{i=1}^{k < n} \mathcal{D}\neg a_i \wedge \neg \mathcal{D}\neg a_{k+1} \rightarrow \mathcal{G}a_{k+1} \quad (\odot\text{-G-gen})$$

An intention is the first mental attitude where the agent compares her outcomes (her preferred courses of action) with the environment she is operating in. Indeed, given an  $\odot$ -chain, it may be the case that neither of the first, say, four outcomes are attainable because facts of the environment (or her own beliefs) state that the opposites of such four outcomes actually hold. (For instance,  $a_1$  may be that the agent would like travel to California, but she believes that her current funds would not cover the whole trip.) Consequently, intentions represent the first level of real commitment of the agent.

Our agents are not omniscient, they may lack the knowledge of certain facts of the world: an agent’s belief represents the fact that she has reasons to accept a

certain statement to be true without having the evidence that it is so. Using her competence/knowledge, she forms an opinion (a belief) of how the environment she is situated in is, or might be. Consequently, she can use her beliefs or the knowledge she possesses about the environment to determine to what outcomes she commits to, depending on her risk attitude. The outcomes she commits to are her intentions.

The axiom hereafter tries to capture many facets of the agent's level of commitment, represented by the three different  $\alpha$ s combined with the four different  $\beta$ s (in fact, the  $\beta_5$  alternative of *wishful thinker* has been introduced for completeness sake only).

$$a_1 \odot \cdots \odot a_n \wedge \left\{ \begin{array}{l} \alpha_1 = \bigwedge_{\substack{i=1 \\ k < n}}^{k < n} \neg a_i \\ \alpha_2 = \bigwedge_{\substack{i=1 \\ k < n}} \mathbf{B} \neg a_i \\ \alpha_3 = \bigwedge_{i=1} \neg a_i \vee \mathbf{B} \neg a_i \end{array} \right\} \wedge \left\{ \begin{array}{l} \beta_1 = \mathbf{B} a_{k+1} \\ \beta_2 = \neg \mathbf{B} \neg a_{k+1} \\ \beta_3 = \mathbf{B} a_{k+1} \vee \neg \mathbf{B} \neg a_{k+1} \\ \beta_4 = a_{k+1} \\ \beta_5 = \neg a_{k+1} \end{array} \right\} \rightarrow \mathbf{I} a_{k+1}. \quad (\text{I})$$

A few comments are in order. In alternative  $\alpha_1$  we have the strongest commitment for the agent (some sort of *omniscient/librarian* agent). In our example, this case model the situation that Alice knows that John is not at home, so she cannot visit him so she will not form the intention to visit him. In here, we use the term omniscient not in the perspective/meaning of a know-it-all agent, but to stress out the fact that the agent does not relies on her own beliefs but she needs to verify the falsity of a statement to proceed in the  $\odot$ -chain about what things really are (she has the burden of proof). This can be view as the most burdensome among our alternatives.

Alternative  $\alpha_2$  is the more 'introspective' one (or more *cautious*): the agent is not interested in how things truly are and her decision process relies only upon what she believes in. It can be described as a more cautious approach because the agent discards a possible outcome based only on her own beliefs, which can be false. For instance, if  $w \models a$  and  $w \models \mathbf{B} \neg a$ , then the agent should derive  $\mathbf{I} a$  (even if an oracle would forecast her eventual failure).

Alternative  $\alpha_3$  can be called as the 'good enough' alternative: if the agent has not the capabilities to verify a certain statement, she 'trusts' in her own beliefs.

Alternative  $\beta_1$  is 'risk-adverse': the agent's decision is based upon on her beliefs only about the truthfulness of the outcome she will try to achieve.

On the contrary, alternative  $\beta_2$  is a 'risk-taking' position, given that the agent looks at whether the opposite outcome holds. Based on her competence/experience, she has evidence that the opposite actually does not hold, and she thus tries to achieve that particular outcome.

The alternative-sequence  $\alpha_3 \wedge \beta_3$  is the most pragmatic one: the agent takes neither a risk-taking, nor a risk-adverse position.

$\beta_4$  represents a non-risk taking position (a *win-win* situation): the agent knows that such an outcome actually holds in the environment she is situated in and then it is feasible to achieve (in our running Alice example, she is already at John's place).

Alternative  $\beta_5$  is a typical case of *wishful thinker* (the agent knows she will fail).

A social intention is an intention that is permitted within the legal system the agent is operating in.

$$\frac{Ia \wedge \neg O\neg a}{SIa} \quad (8)$$

Naturally, in the decision process to decide whether an outcome may be a social intention, the agent uses the  $\odot$ -chains. Therefore, to state that an element in an  $\odot$ -chain is a social intention (say  $a_{k+1}$ ), it must satisfy the following requirements: (1) every element before  $a_{k+1}$  is not an intention or the opposite obligation holds, (2)  $a_{k+1}$  is actually an intention, and (3) the opposite obligation ( $O\neg a_{k+1}$ ) does not hold. This is expressed by the following axiom.

$$a_1 \odot \cdots \odot a_n \wedge \left( \bigwedge_{i=1}^{k < n} \neg Ia_i \vee O\neg a_i \right) \wedge Ia_{k+1} \wedge \neg O\neg a_{k+1} \rightarrow SIa_{k+1}. \quad (SI)$$

To illustrate this axioms consider again Example 1. From (2) we have  $O\neg parking\_downtown$ , and there is no norm to prevent her to take public transports, so  $\neg O\neg public\_transports$ . She does not want to get a parking ticket, so she decided not to take her care, so she does not form the intention to park downtown, thus her next preference is to take public transport, that is  $Ipublic\_transports$ . Thus this intention is also a social intention, i.e., we have  $SIpublic\_transports$ .

In the final part of this paper, we shall prove soundness and completeness results for a system  $S$  containing the schemata presented above.

**Definition 2.** Let  $S$  be a logical system extending the Classical Propositional Calculus (CPC), containing the axiom schemata ( $\otimes$ -contraction), ( $\odot$ -contraction), ( $\otimes$ -shortening), ( $\odot$ -shortening), (**K**), ( $\perp$ - $\Box$ ), (**D**), (**4**), (**5**), (**N**), ( $\otimes$ -**O**), (**O-detachment**), ( $\odot$ -**D**), ( $\odot$ -**G**<sub>1</sub>), ( $\odot$ -**G-gen**), (**I**), and closed under the following rules: ( $\Box$ -RE), ( $\otimes$  and  $\odot$ -RE) and ( $\Box$ -RR).

## 4 Sequence Semantics

Sequence semantics is an extension of neighbourhood semantics. The extension is twofold: (1) we introduce a second neighbourhood like function, and (2) the new function generates a set of sequences of sets of possible worlds instead of set of sets of possible worlds. This extension allows us to provide a clean semantic representation of  $\otimes$ - and  $\odot$ -chains.

The sequence semantics addresses the problem identified in [7] for possible world semantics for deontic logic for compensatory obligations. A compensatory

obligation is a sub-class of a contrary-to-duty obligation, where the violation of the primary obligation is compensated by the fulfilment of the secondary obligation. Compensatory obligations can be modelled by  $\otimes$ -chains. As we have already discussed  $a \otimes b$  means that  $a$  is obligatory, but its violation is compensated by  $b$ , or in other terms it is obligatory to do  $b$  to compensate the violation of the obligation of  $a$ . Thus, a situation where  $a$  does not hold (or  $\neg a$  holds) and  $b$  holds is still deemed as a “legal” situation. Accordingly, when we use a “standard” possible world semantics, there is a deontically accessible world where  $\neg a$  holds, but this implies, according the usual evaluation conditions for permission (something is permitted, if there is a deontically accessible world where it holds), that  $\neg a$  is permitted. However, we have the norm modelling the compensatory obligation that states that  $a$  is obligatory (and if it were not, then there would be no need for  $b$  to compensate for the violation, since, there would be no violation of the obligation of  $a$ ). The sequence semantics solves this problem by establishing that to have an obligation, we must have a norm generating the obligation (where a norm is represented by an  $\otimes$ -chain), and not simply that something is obligatory because it holds in all the deontically accessible worlds. Similarly, when we consider mental-attitudes, an agent forms a mental attitude because the agent has an outcome in mind (where, as we have argued, an outcome is a gradation of alternative objectives).

Before introducing the semantics, we give some technical definitions for operation of *s-zipping*, i.e., the operation that removes repetitions or redundancies occurring in sequences of sets of worlds. This operation is required to capture the intuition described by the  $\otimes$ - and  $\odot$ -contraction axioms.

**Definition 3.** *Given a set of possible worlds  $W$ , let  $X = \langle X_1, \dots, X_n \rangle$  be such that  $X_i \in 2^W$  ( $1 \leq i \leq n$ ). A sequence of sets  $Y$  is s-zipped from  $X$  iff  $Y$  is obtained from  $X$  by applying the following operation: for  $1 \leq k \leq n$ , if  $X_j = X_k$  and  $j < k$ , delete  $X_k$  from the sequence.*

**Definition 4.** *A set  $S$  of sequences of sets of possible worlds is closed downward s-zipping iff if  $X \in S$ , then (i) for all  $Y$  such that  $X$  is s-zipped from  $Y$ ,  $Y \in S$ ; and (ii) for all  $Z$  such that  $Z$  is s-zipped from  $X$ ,  $Z \in S$ .*

Closure under s-zipping essentially determines classes of equivalences for  $\otimes$ -chain and  $\odot$ -chain based on Axioms ( $\otimes$ -**contraction**) and ( $\odot$ -**contraction**).

**Definition 5.** *A sequence frame is a structure*

$$\mathcal{F} = \langle W, \mathcal{C}^O, \mathcal{C}^{\text{out}}, \mathcal{N}^O, \mathcal{N}^B, \mathcal{N}^D, \mathcal{N}^G, \mathcal{N}^I, \mathcal{N}^{\text{Sl}} \rangle,$$

where

- $W$  is a non empty set of possible worlds;
- $\mathcal{C}^O$  and  $\mathcal{C}^{\text{out}}$  are two functions with signature  $W \mapsto 2^{(2^W)^n}$ , such that for every world  $w \in W$ , for every  $X \in \mathcal{C}_w^O$ , and  $Y \in \mathcal{C}_w^{\text{out}}$   $X$  and  $Y$  are closed under s-zipping;
- $\mathcal{N}^O, \mathcal{N}^B, \mathcal{N}^D, \mathcal{N}^G, \mathcal{N}^I$ , and  $\mathcal{N}^{\text{Sl}}$  are functions with Signature  $W \mapsto 2^{2^W}$ .

As we have already said the sequence semantics is an extension and generalisation of neighbourhood semantics. Here the  $\mathcal{N}$  functions are just instances of the standard neighbourhood function that associates sets of propositions (a proposition can be represented by the set of possible world where the proposition holds) to possible worlds. Thus, for example,  $\mathcal{N}_w^{\mathbf{B}}$  denotes the set of propositions that an agent believes at  $w$ . The  $\mathcal{C}$  functions are generalisations of the  $\mathcal{N}$  functions. Instead of a set of propositions, each of these functions associates a set of sequences of propositions to each possible worlds. Each of such sequences describes an order over the propositions in it (with respect to a possible world). Thus,  $\mathcal{C}_w^{\text{out}}$  gives the semantic representation of the preferences of the agents at  $w$ ; similarly,  $\mathcal{C}^{\text{O}}$  models the norms in force in a particular possible world.

**Definition 6.** A sequence model is a structure  $\mathcal{M} = \langle \mathcal{F}, V \rangle$ , where

- $\mathcal{F}$  is a bi-sequence frame, and
- $V$  is a valuation function,  $V: \text{Prop} \mapsto 2^W$

Given a model  $\mathcal{M} = \langle \mathcal{F}, V \rangle$ , let  $\|a\|_V := \{w \mid \models_w^V a\}$ .

**Definition 7.** The valuation function for a sequence model is a follows:

- usual for atoms and boolean conditions,
- $\models_w^V a_1 \otimes \dots \otimes a_n$  iff  $\langle \|a_1\|_V, \dots, \|a_n\|_V \rangle \in \mathcal{C}_w^{\text{O}}$ ,
- $\models_w^V a_1 \odot \dots \odot a_n$  iff  $\langle \|a_1\|_V, \dots, \|a_n\|_V \rangle \in \mathcal{C}_w^{\text{out}}$ ,
- $\models_w^V \Box a$  iff  $\|a\|_V \in \mathcal{N}_w^{\Box}$ , where  $\Box = \{\text{O}, \text{B}, \text{D}, \text{G}, \text{I}, \text{SI}\}$ .

The definition above allows us to characterise a minimal logic satisfying ( $\Box$ -RE), ( $\otimes$  and  $\odot$ -RE), ( $\otimes$ -contraction) and ( $\odot$ -contraction) [8].

**Definition 8.** A neighbourhood function  $\mathcal{N}$  is

- supplemented if  $X \cap Y \in \mathcal{N}$ , then  $X \in \mathcal{N}$  and  $Y \in \mathcal{N}$ ;
- closed under the intersection if  $X \in \mathcal{N}$  and  $Y \in \mathcal{N}$ , then  $X \cap Y \in \mathcal{N}$ ;
- contains the unit if  $W \in \mathcal{N}$ .

A neighbourhood function is a filter if it is supplemented, closed under intersection and contains the unit. A neighbourhood function is a quasi-filter if it is supplemented and closed under intersection.

The definition below gives the conditions of the frames for the various axioms.

**Definition 9.** A sequence model for  $\mathbf{S}$ , as defined in 2, is a structure  $\mathcal{M} = \langle \mathcal{F}, v \rangle$  satisfying the following conditions:

- (**K**) and (**N**):  $\mathcal{N}_w^{\mathbf{B}}$  is a filter.
- ( $\perp$ - $\Box$ ):  $\emptyset \notin \mathcal{N}_w^{\Box}$ , for  $\Box \in \{\text{O}, \text{B}, \text{D}, \text{G}, \text{I}, \text{SI}\}$ .
- (**D**): If  $X \in \mathcal{N}_w^{\Box}$ , then  $\neg X \notin \mathcal{N}_w^{\Box}$ , for  $\Box \in \{\text{O}, \text{B}, \text{G}, \text{I}, \text{SI}\}$ .
- (**4**): If  $X \in \mathcal{N}_w^{\mathbf{B}}$ , then  $\{x \in W : X \in \mathcal{N}_x^{\mathbf{B}} \in \mathcal{N}_w^{\mathbf{B}}\}$ .
- (**5**): If  $X \notin \mathcal{N}_w^{\mathbf{B}}$ , then  $\{x \in W : X \notin \mathcal{N}_x^{\mathbf{B}} \in \mathcal{N}_w^{\mathbf{B}}\}$ .
- ( $\Box$ -RR):  $\mathcal{N}_w^{\Box}$ , for  $\Box \in \{\text{O}, \text{D}, \text{G}, \text{I}, \text{SI}\}$ , is a quasi-filter.

- ( $\otimes$ -shortening) and ( $\odot$ -shortening): If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^x$ , then, for  $x \in \{\mathbf{O}, \text{out}\}$ ,  $\langle X_1, \dots, X_{n-1} \rangle \in \mathcal{C}_w^x$ .
- ( $\odot$ -O): If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\mathbf{O}}$ , then  $X_1 \in \mathcal{N}_w^{\mathbf{O}}$ .
- (O-detachment) If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\mathbf{O}}$  and  $w \notin X_i$  for  $1 \leq i \leq k$  and  $k < n$ , then  $X_{k+1} \in \mathcal{N}_w^{\mathbf{O}}$ .
- ( $\odot$ -D): If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$ , then  $X_i \in \mathcal{N}_w^{\mathbf{D}}$ , with  $1 \leq i \leq n$ .
- ( $\odot$ -G<sub>1</sub>): If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and  $-X_1 \notin \mathcal{N}_w^{\mathbf{D}}$ ,  $X_1 \in \mathcal{N}_w^{\mathbf{G}}$ .
- ( $\odot$ -G-gen): If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and, for  $1 \leq i \leq k$  and  $k < n$ ,  $-X_i \in \mathcal{N}_w^{\mathbf{D}}$  and  $-X_{k+1} \notin \mathcal{N}_w^{\mathbf{D}}$ , then  $X_{k+1} \in \mathcal{N}_w^{\mathbf{G}}$ .
- (I): If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and, for  $1 \leq i \leq k$  and  $k < n$ ,  
either  
 $\alpha_1: w \notin X_i$ ;  
 $\alpha_2: -X_i \in \mathcal{N}^{\mathbf{B}}$ ;  
 $\alpha_3: w \notin X_i$  or  $-X_i \in \mathcal{N}^{\mathbf{B}}$   
and either  
 $\beta_1: X_{k+1} \in \mathcal{N}^{\mathbf{B}}$ ;  
 $\beta_2: -X_{k+1} \notin \mathcal{N}^{\mathbf{B}}$ ;  
 $\beta_3: X_{k+1} \in \mathcal{N}^{\mathbf{B}}$  or  $-X_{k+1} \notin \mathcal{N}^{\mathbf{B}}$ ;  
 $\beta_4: w \in X_{k+1}$ ;  
 $\beta_5: w \notin X_{k+1}$ ,  
then  $X_{k+1} \in \mathcal{N}_w^{\mathbf{I}}$ .
- (SI): If  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and for  $1 \leq i \leq k$  and  $k < n$ ,  $X_i \notin \mathcal{N}_w^{\mathbf{I}}$  or  $-X_i \in \mathcal{N}_w^{\mathbf{O}}$ ,  $X_{k+1} \in \mathcal{N}_w^{\mathbf{I}}$ , and  $-X_{k+1} \notin \mathcal{N}_w^{\mathbf{O}}$ , then  $X_{k+1} \in \mathcal{N}_w^{\text{SI}}$ .

A few comments are in order. The conditions for axioms (**K**), (**D**), (**4**), (**5**), (**N**) and ( $\perp$ - $\square$ ) and the inference rule ( $\square$ -RR) are the standard conditions for such axioms in neighbourhood semantics (see [4]). The conditions for the remaining axioms exploit the strong correspondence between propositions and truth sets that allows for a semantic rewriting of the axioms. Axioms ( $\otimes$ -shortening), ( $\odot$ -shortening), ( $\otimes$ -O) and (O-detachment) were first proposed by [8] for the use of the sequence semantics for logics to model norms and obligations.

## 5 Soundness and Completeness

In this section we study the soundness and completeness of the logics defined in Sect. 3. Completeness is based on adaptation of the standard Lindenbaum's construction for modal (deontic) neighbourhood semantics (see [4]).

**Definition 10 ( $\mathcal{L}$ -maximality).** A set  $w$  is  $\mathcal{L}$ -maximal iff for any formula  $a$  of  $\mathcal{L}$ , either  $a \in w$ , or  $\neg a \in w$ .

**Lemma 1 (Lindenbaum's Lemma).** Any  $\mathcal{S}$ -consistent set  $w$  of formulae in the language  $\mathcal{L}$  can be extended to a  $\mathcal{S}$ -consistent  $\mathcal{L}$ -maximal set  $w^+$ .

*Proof.* Let  $a_1, a_2, \dots$  be an enumeration of all the possible formulae in  $\mathcal{L}$ .

–  $w_0 := w$ ;



- $w_{n+1} = w_n \cup \{a_n\}$  if its closure under the axioms and rules of  $\mathbf{S}$  is consistent,
- $w \cup \{\neg a_n\}$  otherwise;
- $w^+ = \bigcup_{n \geq 0} w_n$ .

The construction of a sequence canonical model is as follows.

**Definition 11 (S-Canonical Models).** *A sequence canonical model*

$$\mathcal{M} = \langle W, \mathcal{C}^{\mathcal{O}}, \mathcal{C}^{\text{out}}, \mathcal{N}^{\mathcal{O}}, \mathcal{N}^{\mathcal{B}}, \mathcal{N}^{\mathcal{D}}, \mathcal{N}^{\mathcal{G}}, \mathcal{N}^{\mathcal{I}}, \mathcal{N}^{\text{SI}}, V \rangle$$

for the system  $\mathbf{S}$  is defined as follows:

1.  $W$  is the set of all the  $\mathcal{L}$ -maximal consistent sets.
2. For any propositional letter  $p \in \text{Prop}$ ,  $\|p\|_V := \|p\|_{\mathcal{L}}$ , where

$$\|p\|_{\mathcal{L}} := \{w \in W \mid p \in w\}.$$

3. Let  $\mathcal{C}^{\mathcal{O}} := \bigcup_{w \in W} \mathcal{C}_w^{\mathcal{O}}$ , where, for each  $w \in W$ ,

$$\mathcal{C}_w^{\mathcal{O}} := \{ \langle \|a_1\|_V, \dots, \|a_n\|_V \rangle \mid \bigotimes_{i=1}^n a_i \in w \},$$

where each  $a_i$  is a meta-variable for a Boolean formula.

4. Let  $\mathcal{C}^{\text{out}} := \bigcup_{w \in W} \mathcal{C}_w^{\text{out}}$ , where, for each  $w \in W$ ,

$$\mathcal{C}_w^{\text{out}} := \{ \langle \|a_1\|_V, \dots, \|a_n\|_V \rangle \mid \bigodot_{i=1}^n a_i \in w \},$$

where each  $a_i$  is a meta-variable for a Boolean formula.

5. Let  $\mathcal{N}^{\square} := \bigcup_{w \in W} \mathcal{N}_w^{\square}$  where for each world  $w$ ,

$$\mathcal{N}_w^{\square} := \{ \|a_i\|_V \mid \square a_i \in w \},$$

where  $\square \in \{\mathcal{O}, \mathcal{B}, \mathcal{D}, \mathcal{G}, \mathcal{I}, \text{SI}\}$ .

**Lemma 2 (Truth Lemma for Canonical Sequence Models).** *If  $\mathcal{M} = \langle W, \mathcal{C}^{\mathcal{O}}, \mathcal{C}^{\text{out}}, \mathcal{N}^{\mathcal{O}}, \mathcal{N}^{\mathcal{B}}, \mathcal{N}^{\mathcal{D}}, \mathcal{N}^{\mathcal{G}}, \mathcal{N}^{\mathcal{I}}, \mathcal{N}^{\text{SI}}, V \rangle$  is canonical for  $\mathbf{S}$ , then for any  $w \in W$  and for any formula  $A$ ,  $A \in w$  iff  $\models_w^V A$ .*

*Proof.* Given the construction of the canonical model, this proof is easy and can be given by induction on the length of an expression  $A$ . We consider only some relevant cases.

Assume  $A$  has the form  $a_1 \otimes \dots \otimes a_n$ . If  $A \in w$ , by definition of canonical model, then there is a sequence  $\langle \|a_1\|_V, \dots, \|a_n\|_V \rangle \in \mathcal{C}_w^{\mathcal{O}}$ . Following from the semantic clauses given to evaluate  $\otimes$ -formulae, it holds that  $\models_w^V a_1 \otimes \dots \otimes a_n$ . For the opposite direction, assume that  $\models_w^V a_1 \otimes \dots \otimes a_n$ . By definition, there is  $\mathcal{C}_w$  which contains an ordered  $j$ -tuple  $\langle \|a_1\|_V, \dots, \|a_n\|_V \rangle$  and by construction  $a_1 \otimes \dots \otimes a_n \in w$ . Clearly the same argument holds in the case of operator  $\odot$ .

If, on the other hand,  $A$  has the form  $\square b$ , where  $\square \in \{\mathcal{O}, \mathcal{B}, \mathcal{D}, \mathcal{G}, \mathcal{I}, \text{SI}\}$ , and  $\square b \in w$ , then  $\|b\|_V \in \mathcal{N}_w^{\square}$  by construction, and by definition  $\models_w^V \square b$ . Conversely, if  $\models_w^V \square b$ , then  $\|b\|_V \in \mathcal{N}_w^{\square}$  and, by construction of  $\mathcal{N}^{\square}$ ,  $\square b \in w$ .

It is easy to verify that the canonical model exists, it is not empty, and it is a sequence semantics model.

Consider any formula  $A \notin \mathcal{S}$ ;  $\{\neg A\}$  is consistent and it can be extended to a maximal set  $w$  such that for some canonical model,  $w \in W$ . By Lemma 2,  $\not\models_w^V A$ .

**Corollary 1.** *The system  $\mathcal{S}$  is sound and complete with respect to the class of bi-sequence frames.*

**Lemma 3.** *The canonical model for  $\mathcal{S}$  enjoys all the properties listed in Definition 9.*

*Proof.* For the cases for  $(\Box\text{-RE})$ ,  $(\mathbf{K})$ ,  $(\mathbf{N})$ ,  $(\perp\text{-}\Box)$ ,  $(\mathbf{D})$ ,  $(\mathbf{4})$  and  $(\mathbf{5})$ , see [4]. The proofs for the cases for  $(\otimes\text{ and } \odot\text{-RE})$ ,  $(\otimes\text{-contraction})$ ,  $(\odot\text{-contraction})$ ,  $(\otimes\text{-shortening})$ ,  $(\odot\text{-shortening})$ ,  $(\otimes\text{-O})$  and  $(\mathbf{O}\text{-detachment})$  are given in [8].

The proof of the remaining cases is rather straightforward and it follows the structure of the axioms involved.

- $(\odot\text{-D})$  Assume  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$ . Then, by construction of the canonical model, for  $1 \leq i \leq n$ , it holds that  $X_i = \|a_i\|_V$  and  $a_1 \odot \dots \odot a_n \in w$ . Thus  $\mathbf{D}a_i \in w$  for  $1 \leq i \leq n$  by  $(\odot\text{-D})$ , hence  $X_i \in \mathcal{N}_w^{\mathbf{D}}$ , with  $1 \leq i \leq n$ .
- $(\odot\text{-G}_1)$  Assume  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and  $\neg X_1 \notin \mathcal{N}_w^{\mathbf{D}}$ . Then, by construction,  $a_1 \odot \dots \odot a_n \wedge \neg \mathbf{D}\neg a_1 \in w$ , where for  $1 \leq i \leq n$ ,  $X_i = \|a_i\|_V$ . By  $(\odot\text{-G}_1)$ ,  $\mathbf{G}a_1 \in w$  and  $X_1 \in \mathcal{N}_w^{\mathbf{G}}$ .
- $(\odot\text{-G-gen})$  Suppose  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and, for  $1 \leq i \leq k$  and  $k < n$ ,  $\neg X_i \in \mathcal{N}_w^{\mathbf{D}}$  and  $\neg X_{k+1} \notin \mathcal{N}_w^{\mathbf{D}}$ , then  $a_1 \odot \dots \odot a_n \wedge \bigwedge_{i=1}^{k < n} \mathbf{D}\neg a_i \wedge \neg \mathbf{D}\neg a_{k+1} \in w$  by construction of the canonical model. By  $(\odot\text{-G-gen})$ ,  $\mathbf{G}a_{k+1} \in w$  and  $X_{k+1} \in \mathcal{N}_w^{\mathbf{G}}$ .
- (I) Assume  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and, for  $1 \leq i \leq k$  and  $k < n$ ,  $\alpha_1: w \notin X_i$  and  $\beta_1: X_{k+1} \in \mathcal{N}^{\mathbf{B}}$ . Then  $a_1 \odot \dots \odot a_n \wedge \bigwedge_{i=1}^{k < n} \neg a_i \wedge \mathbf{B}a_{k+1} \in w$ . Hence, by (I),  $\mathbf{I}a_{k+1} \in w$  and  $X_{k+1} \in \mathcal{N}_w^{\mathbf{I}}$ . The proof for the other cases is similar.
- (SI) Suppose  $\langle X_1, \dots, X_n \rangle \in \mathcal{C}_w^{\text{out}}$  and for  $1 \leq i \leq k$  and  $k < n$ ,  $X_i \notin \mathcal{N}_w^{\mathbf{I}}$  or  $\neg X_i \in \mathcal{N}_w^{\mathbf{O}}$ ,  $X_{k+1} \in \mathcal{N}_w^{\mathbf{I}}$ , and  $\neg X_{k+1} \notin \mathcal{N}_w^{\mathbf{O}}$ . Then, by construction of the canonical model,  $a_1 \odot \dots \odot a_n \wedge (\bigwedge_{i=1}^{k < n} \neg \mathbf{I}a_i \vee \mathbf{O}\neg a_i) \wedge \mathbf{I}a_{k+1} \wedge \neg \mathbf{O}\neg a_{k+1} \in w$ . By (SI) and modus ponens,  $\mathbf{S}Ia_{k+1} \in w$  and thus  $X_{k+1} \in \mathcal{N}_w^{\mathbf{SI}}$ .

## 6 Conclusions

This paper offered a semantic study of the  $\otimes$  and  $\odot$  operators originally introduced in [5] to model deontic reasoning and contrary-to-duty obligations. We showed that a suitable axiomatisation was able to capture characteristics in the context of multi-modal logics the unified framework by [9] for agents' motivational and deliberative components where goals, desires, and intentions are *different facets of the same phenomenon*, all of them being goal-like attitudes. In particular, we proved that  $\otimes$ - and  $\odot$  expressions can be characterised in a class of structures extending neighbourhood frames with sequences of sets of worlds.

We argued that both the formalism, and the semantics can be employed, with some adjustments, to grasp various forms of reasoning about BDI-like agents.

A number of open research issues are left for future work. The logic of [9] investigates how to characterise different degrees and types of goal-like mental attitudes of agents. These works assume defeasible logic as the underlying logic and they are restricted to literals. However, they show that the extension of defeasible logic obtained from adding  $\otimes$  and  $\odot$  are still computationally feasible. The natural question is to see how to use the sequence semantics we have presented in this paper to capture the different intuitions of  $\otimes$  and  $\odot$  discussed in the above mentioned work. In addition we plan to explore decidability questions using, for example, the filtration methods. The fact that neighbourhoods contain sequences of sets of worlds instead of sets is not expected to make the task significantly harder than the one in standard neighbourhood semantics for modal logics.

Second, we expect to enrich the language and allow for nesting of  $\otimes$ - and  $\odot$ -expressions, thus having formulae like  $a \otimes \neg(b \otimes c) \otimes d$ . We argued in [5] that the meaning of those formulae is not clear in deontic reasoning. However, a semantic analysis of them in the sequence semantics can clarify the issue. Indeed, in the current language we can evaluate in any world  $w$  formulae like  $\neg(a \otimes b)$ .

## References

1. Andrighetto, G., Governatori, G., Noriega, P., van der Torre, L.W.N. (eds.) Normative Multi-agent Systems. Dagstuhl Follow-Ups, vol. 4. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2013)
2. Brewka, G., Benferhat, S., Le Berre, D.: Qualitative choice logic. *Artif. Intell.* **157**(1-2), 203–237 (2004)
3. Broersen, J., Dastani, M., Hulstijn, J., van der Torre, L.: Goal generation in the BOID architecture. *Cogn. Sci. Q.* **2**(3-4), 428–447 (2002)
4. Chellas, B.F.: *Modal Logic, An Introduction*. Cambridge University Press, Cambridge (1980)
5. Governatori, G., Rotolo, A.: Logic of violations: a Gentzen system for reasoning with contrary-to-duty obligations. *Australas. J. Logic* **4**, 193–215 (2006)
6. Governatori, G., Rotolo, A.: BIO logical agents: norms, beliefs, intentions in defeasible logic. *Auton. Agent. Multi-agent Syst.* **17**(1), 36–69 (2008)
7. Governatori, G.: Thou shalt is not you will. In: Atkinson, K. (ed.) *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Law*, pp. 63–68. ACM, New York (2015)
8. Governatori, G., Olivieri, F., Calardo, E., Rotolo, A.: Sequence semantics for norms and obligations. In: *Proceedings Deontic Logic and Normative Systems*, pp. 93–108. College Publications (2016)
9. Governatori, G., Olivieri, F., Scannapieco, S., Rotolo, A., Cristani, M.: The rational behind the concept of goal. *Theory Pract. Logic Program.* **16**(3), 296–324 (2016)
10. Thomason, R.H.: Desires, defaults: a framework for planning with inferred goals. In: Cohn, A.G., Giunchiglia, F., Selman, B. (eds.) *KR2000*. Morgan Kaufmann, San Francisco (2000)