

On the Axiomatization of Elgesem’s Logic of Agency and Ability

Guido Governatori
School of ITEE, The University of Queensland,
Brisbane, QLD 4072, Australia
email `guido@itee.uq.edu.au`

Antonino Rotolo
CIRSFID, University of Bologna
Via Galliera 3, 40121, Bologna, Italy
email `rotolo@cirfid.unibo.it`

1 Introduction

Modal logic of agency is a traditional research field in philosophical logic. Roughly speaking, the approach adopts the general policy to abstract from making explicit state changes and from considering the temporal dimension in describing actions. In fact, actions are simply taken to be relationships between agents and states of affairs. Thus, the conceptual qualification of these relations is made by using suitable modal operators to represent, for example, that an agent “brings it about” or “sees to it” that A , or that such agent is “able” to realise A , or again that she “attempts” to achieve it.

It is well known that modal logic of agency has a number of drawbacks. As recently summarised in [24], the main limit of this approach, as found in the literature, is that it is “too abstract”. For example, it is not usually captured the difference between the modal qualifications “sees to it” and “brings it about”. Both expressions are in general represented by a single modal operator, despite the fact that the former exhibits a clear intentional character, whereas the latter may refer as well to unintentional actions [11]. Secondly, for the purpose of analysing the structure of multi-agent contexts it is crucial to distinguish between direct actions and indirect actions. This is necessary, for example, to account for the notions of influence and control of an agent over other agents [14, 15, 19, 20]. While these problems may be, or have been, solved by providing suitable integrations and new operators within the same paradigm of modal logic of agency, a last drawback is inherent in the paradigm as such. In fact, “sometimes it is essential to be able to refer to the *means* by which an agent brings about a state of affairs”, as for example by referring to specific actions performed to achieve a goal [24]. As it is well-known, this shows that modal logic of agency is less expressive than other formal theories of action, such as dynamic logics. On the other hand, this last limit is also an advantage. Although the abstractness of modal logic of agency does not make the language very expressive in itself, it allows flexibility for the easy combination of agency with a number of other concepts, such as powers,

obligations, beliefs, etc, in a multi-modal setting. This perhaps explains why the approach has been recently used to analyse some crucial aspects of normative and institutional domains (see, e.g., [3, 9, 14, 15, 19, 20]).

The formal properties of modal logic of agency have been extensively investigated, and a number of variants and axiomatizations can be found in literature (see, e.g., [1, 2, 4, 6, 7, 13, 17–22]). Despite this great variety, it is possible to identify a minimal core of axioms that seem to characterize indisputably some aspects of agency. The recent contributions by Dag Elgesem are meant to work in this direction [6, 7]. We will focus here on two praxeological notions among those considered by him¹.

The first is the idea of personal and direct action to realise a state of affairs. In the mentioned general view, this idea is formalised by the well-known modal operator E , such that a formula like E_iA means that the agent i brings it about that A . Elgesem’s logic of E is a classical non-normal system [5], namely is closed under logical equivalence, and is characterized by the following schemas.

$$E_iA \rightarrow A \tag{1}$$

(1) is recognised as valid by almost all theories of agency. It is nothing but the usual axiom T of modal logic, and it expresses the successfulness of actions which is behind the common reading of “bring about” concept.

$$\neg E_i\top \tag{2}$$

The axiom (2), also named No, is used to capture the very concept of agency at hand, according to which the occurrence of any state of affairs, in the scope of E_i , is the result of an action of i . In other words, if i had not behaved in the way she did, the world might have been different. In this perspective, at least no agent can bring about what is logically unavoidable.

$$(E_iA \wedge E_iB) \rightarrow E_i(A \wedge B) \tag{3}$$

This third schema, C or Agglomeration, follows from the co-temporality of actions implicitly assumed within the paradigm of modal logic of agency. In fact, if the agent i realises A and B , presumably by performing two distinct actions, it can be also said that i brings it about that $A \wedge B$ only if the two actions have been performed at the same time. As it is argued by Elgesem, however, the converse of 3 must be rejected because, in presence of it, substitution of equivalents (i) plus 2 make the logic inconsistent whenever at least one action is performed, (ii) gives the usual rule RM ($\vdash A \rightarrow B / \vdash \Box A \rightarrow \Box B$), which is not acceptable in the logic for E .

The second praxeological concept, analysed by Elgesem and considered here, is agents’ “practical ability” to realise states of affairs. This praxeological qualification is represented by the modal operator C . Accordingly, C_iA expresses that i is capable of realising A . The logic for C is quite weak. It is closed as well under logical equivalence and is characterised by the following principles.

$$E_iA \rightarrow C_iA \tag{4}$$

¹As we will see in a few moments, the two concepts are those of “bring about” and “practical ability”. Elgesem formalises them as Does and Ability respectively, such that both operators are, as expected, indexed by agents. For the sake of simplicity, we will adopt a different notation, which is quite common in the literature (see, e.g., [15]). Thus the first is represented by the operator E , while the second by C . Of course, both are labelled by agents as well.

This schema states a strong connection between ability and agency. Of course, the latter implies the former, in presence of axiom (1): If i realises successfully A , it is obvious that i is able to do this.

$$\neg C_i \top \quad (5)$$

This last axiom is the natural counterpart of schema (2) for E . As we have alluded to, both express jointly the idea of avoidability, namely that the occurrence of a state of affairs cannot be caused by one agent if the goal obtains in every state of the world².

In the next sections we will analyse some aspects of Elgesem's semantics for the above operators. The focus will be then on a decisive, but quite solvable, problem arising from his own semantic characterisation of the logic of agency and ability.

2 An Axiomatization for Agency and Ability

Elgesem's analysis starts from semantical considerations [6, 7]. His aim is to give a fresh account of Sommerhoff's theory of goal-directness. The semantics is given in terms of selection function models, where a selection function model \mathcal{E} is a structure $\langle W, f, v \rangle$ where W is a (non empty) set of possible worlds, f is a selection function from $\mathcal{P}(W) \times W$ to $\mathcal{P}(W)$, and v assigns to each propositional letter a subset of W .³

Each formula corresponds to a set of worlds, the set of worlds where it is true, and a world describes the formulas true at it; thus a formula corresponds to a state of affairs, and it determines all worlds where the state of affairs is true. The selection function identifies then the worlds relative to the actual world where a goal (state of affairs) has been realized.

For convenience, before providing the valuation clauses for the formulas, we define the notion of truth set, i.e., the set of worlds where a formula is true.

DEFINITION 1. The truth set of a formula A wrt to a model \mathcal{M} , $\|A\|^{\mathcal{M}}$ is thus defined:

$$\|A\|^{\mathcal{M}} = \{w \in W : w \models_{\mathcal{M}} A\}.$$

An Elgesem model is a selection function model \mathcal{E} satisfying the following valuation clauses:

²According to Elgesem, the full idea of avoidability requires to focus on two different, but interconnected, aspects. The first corresponds to the negative conditions stated by (2) and (5). Both schemas are aimed to state that no agent brings about logical truths. The second claim is that "an agent's behaviour, when he brings about something, is instrumental in the production of that which he brings about". This general idea corresponds to saying, positively and with respect to any state of affairs A , that "if the agent had not behaved in the way he did when he brought it about that A , then he might not have brought it about that A ". The last requirement is rendered by defining suitable dyadic operators and principles which reflect Elgesem's own philosophical interpretation of agency [7]. This second aspect will not be considered here, since it does not seem relevant with regard to the aims of this paper.

³Elgesem's semantics for the modal logic of agency and ability is a structure $\langle W, f_1, \dots, f_n, V \rangle$ (cf. [7, p. 20] and [6, p. 54]), where each f_i , $1 \leq i \leq n$ is a function as in the structure described above and i is an agent. Since there are no interactions among the agents and all functions f_i are independent from each other and obey the same conditions, we can restrict ourselves to the case of a single agent. Elgesem also considers some foundational aspects of the notions he deals with and introduces some additional functions in order to capture the idea of avoidability and accident. However those functions do not play any relevant role in the characterisation of the modal operators E and C . The valuation function and the constraints on the model are given in terms of properties of f . The other functions are used to specify constraints on concrete instances of f . Finally V is a valuation function while v is an assignment.

- S1. $w \models_{\mathcal{E}} p$ iff $w \in v(p)$;
- S2. $w \models_{\mathcal{E}} \neg A$ iff $w \not\models_{\mathcal{E}} A$;
- S3. $w \models_{\mathcal{E}} A \rightarrow B$ iff $w \not\models_{\mathcal{E}} A$ or $w \models_{\mathcal{E}} B$;
- S4. $w \models_{\mathcal{E}} EA$ iff $w \in f(\|A\|^{\mathcal{E}}, w)$;
- S5. $w \models_{\mathcal{E}} CA$ iff $f(\|A\|^{\mathcal{E}}, w) \neq \emptyset$.⁴

The notion of truth in a model and validity are defined as usual.

It is immediate to see that (S4) and (S5) together imply the validity of (4), namely $EA \rightarrow CA$. Notice that Elgesem uses only one selection function to represent the two modal operators E and C . This is crucial in his philosophical approach to agency because $f(\|A\|, w)$ corresponds exactly to the set of worlds where an agent realizes her ability, relative to the actual world w , to bring about the goal A . In this perspective, ability and agency are two facets of the same general concept.

Then Elgesem goes on and discusses the conditions required to characterise the modal operators of agency (E) and ability (C); though the two operators are defined by the same selection function, he treats them as independent operators (even if C corresponds to the possibility operator of E , they are not duals, and cannot be defined in terms of each other in the present setting).

To characterise the other principles Elgesem imposes the following conditions on the selection function f :

- E1 $f(W, w) = \emptyset$;
- E2 $f(X, w) \cap f(Y, w) \subseteq f(X \cap Y, w)$.
- E3 $f(X, w) \subseteq X$;

Condition E1 says that a goal that is realized in every world is not a state the agent is able to bring about. As an immediate consequence of this constraints we have the validity of (5) and (2).

Condition E2, corresponding to the agglomeration principle for E (3), is motivated by the idea that the ability needed for the intersection of A and B is not more general than the ability to do A and the ability to do B .

Finally E3 makes explicit the idea that in all worlds where an agent realizes her ability to bring about a goal, the goal is indeed realized. It is easy to see that it validates the success principle (1).

To sum up, let us recall synoptically Elgesem's axiomatization for the logic of agency and ability (let us call the resulting logic \mathcal{L}_1).

A0 propositional logic

A1 $\neg C\top$,

⁴From now on, whenever clear from the context we drop subscripts and superscripts.

A2 $EA \wedge EB \rightarrow E(A \wedge B)$,

A3 $EA \rightarrow A$,

A4 $EA \rightarrow CA$;

plus Modus Ponens and

$$\frac{A \equiv B}{EA \equiv EB} RE_E \quad \frac{A \equiv B}{CA \equiv CB} RE_C \quad (6)$$

As we have seen Elgesem also considers the principle $\neg E\top$; however this principle is redundant since it can be easily derived from A1 and the contrapositive of A4.

Another interesting principle, which can be derived from the success axiom for the operator E (A3) is $\neg E\perp$. This principle states that nobody can realize an inconsistent (impossible) state. But, what about the corresponding principle that nobody is capable to produce an inconsistent state?

$$\neg C\perp \quad (7)$$

This principle is valid in the proposed selection function semantics, but, as we shall see in the next section, is not provable in \mathcal{L}_1 .

Let \mathcal{E} be an Elgesem model. For every world w in \mathcal{E} we have

$$w \models_{\mathcal{E}} \neg C\perp \iff w \not\models_{\mathcal{E}} C\perp \iff f(\|\perp\|^{\mathcal{E}}, w) = \emptyset.$$

According to condition E3 $\forall w \in W, f(X, w) \subseteq X$ and, $\|\perp\|^{\mathcal{E}} = \emptyset$; hence $f(\|\perp\|^{\mathcal{E}}, w) \subseteq \|\perp\|^{\mathcal{E}} = \emptyset$. According to the intended interpretation $\neg C\perp$ means that an agent is not able to realize the impossible (here with impossible we understand an inconsistent state of affairs). This reading seems appropriate in a physical (practical) conception of the notion of ability. However there are other interpretations where such condition might be relaxed. For example Hintikka [12] proposes a reading where impossible worlds are worlds where we have a partial knowledge of the structure of the world and some contradictions do not appear to be as such, unless we perform a deeper analysis of them. A second interpretation where $C\perp$ can be accepted is when we have a “normative” reading of C . As we have alluded to, in the last few years logics of agency and ability have been used, in conjunction with deontic logics, to model the relationships among (autonomous) agents in agent societies conceived as normative systems [3, 9, 14, 15, 20]. In this interpretation we can use C to describe, among other types of “normative” actions, the ability of an agent to create a new normative position. In this perspective, it is indeed possible for a law-maker to draft an inconsistent norm. However, its inconsistency could prevent it from being a true “legal” norm, at least in the event it is accepted the view that all norms must be logically compliant with (and also violable). Of course, a lot depends on the exact axiomatization adopted for praxeological as well as for deontic notions. If O stands as usual for the deontic operator of obligation, it may be argued, for example, that $O\perp$ (and/or $O\top$) are meaningless and so self-contradictory (this view is adopted, for example, in [15]; see also [23]). But nothing in theory is against accepting that an agent may be logically able to issue bizarre norms like $O\perp$, for the simple reason that any logic for the operator O does not include the axiom T. The expression $C_i O\perp$ may be thus accepted. Perhaps, the problem at stake here is that the right way to approach these questions requires to focus on the *normative power* to issue norms, rather than on the practical ability to do this. But, of course, this is outside the scope of the paper.

3 Neighbourhood Models

As we have seen in the previous section $\neg C\perp$ is valid, but we have alluded that it is not provable from \mathcal{L}_1 , hence \mathcal{L}_1 is incomplete wrt the intended semantics. To show that \mathcal{L}_1 is incomplete wrt \mathcal{E} we have to provide a class of models such that \mathcal{L}_1 is complete for it and $\neg C\perp$ is false. While it is possible to devise a class of selection function models for \mathcal{L}_1 (see Section 4) we prefer to introduce models with a different structure.⁵

A neighbourhood model \mathcal{N} is a structure $\langle W, N^C, N^E, v \rangle$ where W is a set of possible worlds, N^C and N^E are functions from W to $\mathcal{P}(\mathcal{P}(W))$, and v assign subset of W to atomic letters.

The valuation clauses for atomic and boolean formulas are as usual while those for modal operators are given below.

DEFINITION 2. Let w be a world in $\mathcal{N} = \langle W, N^C, N^E, v \rangle$:

N1 $w \models_{\mathcal{N}} CA$ iff $\|A\|^{\mathcal{N}} \in N_w^C$;

N2 $w \models_{\mathcal{N}} EA$ iff $\|A\|^{\mathcal{N}} \in N_w^E$.

It is natural to add some conditions on the functions N in neighbourhood models to validate the axioms A1–A4.

C1 $W \notin N_w^C$;

C2 if $X \in N_w^E$ and $Y \in N_w^E$ then $X \cap Y \in N_w^E$;

C3 if $X \in N_w^E$ then $w \in X$;

C4 $N_w^E \subseteq N_w^C$.

THEOREM 3. $\vdash_{\mathcal{L}_1} A$ iff $\models_{\mathcal{N}} A$.

Proof. The proof is a straightforward extension of that given in [5] using minimal canonical models. □

It is easy to provide a neighbourhood model that falsifies $\neg C\perp$. Let $W = \{w\}$, $N_w^E = \emptyset$ and $N_w^C = \{\emptyset\}$. Here, $\|\perp\| = \emptyset \in N_w^C$, therefore $w \models C\perp$ and $w \not\models \neg C\perp$. Hence we have the following result:

PROPOSITION 4. $\not\models_{\mathcal{L}_1} \neg C\perp$.

An immediate consequence of Proposition 4 is that \mathcal{L}_1 is incomplete with respect to the intended selection function semantics \mathcal{E} . It is possible, however, to regain completeness by adding $\neg C\perp$ as axiom to \mathcal{L}_1 (let us call the resulting logic \mathcal{L}_2).

PROPOSITION 5. Let $\mathcal{N}' = \langle W, N^E, N^C, v \rangle$ a neighbourhood model and $\mathcal{E} = \langle W, f, v \rangle$ be an Elgesem model satisfying the following conditions:

⁵As we shall see the difference between the two types of semantics is just in the intuition behind them; in fact, mathematically, they are equivalent and both neighbourhood semantics and selection function semantics are also known as Scott-Montague semantics (cf. [10]).

1. $w \in f(\|A\|^\mathcal{E}, w)$ iff $\|A\|^{\mathcal{N}'} \in N_w^E$; and
2. $f(\|A\|^\mathcal{E}, w) \neq \emptyset$ and $\|A\|^\mathcal{E} \neq W$ iff $\|A\|^{\mathcal{N}'} \in N_w^C$.⁶

Then $\models_{\mathcal{E}} A$ iff $\models_{\mathcal{N}'} A$.

Moreover \mathcal{E} satisfies conditions E1, E2 and E3 iff \mathcal{N}' satisfies conditions C1–C4, and $\emptyset \notin N_w^C$, for every $w \in W$.

The above proposition shows that any selection function models can be transformed in an equivalent neighbourhood models. However such models must satisfy the condition

$$C5 \quad \forall w, \emptyset \notin N_w^C,$$

which is known to correspond to the axiom $\neg C\perp$. Hence we have the following theorem.

THEOREM 6.

1. $\vdash_{\mathcal{L}_2} A$ iff $\models_{\mathcal{N}'} A$;
2. $\vdash_{\mathcal{L}_2} A$ iff $\models_{\mathcal{E}} A$.

The above theorem proves that \mathcal{E} does not determine \mathcal{L}_1 but \mathcal{L}_2 (i.e., $\mathcal{L}_1 + \neg C\perp$). In the next section we will investigate whether there is a class of selection function models that characterises \mathcal{L}_1 .

4 Completeness Regained

In the previous section we have seen that it is possible to regain completeness by using neighbourhood semantics with two neighbourhood functions, one for C (N^C) and one for E (N^E) plus the condition that N^E is included in N^C . Obviously, by the well-known equivalence between selection function semantics and neighbourhood semantics [10], we can use a semantics with two selection functions; but what about a selection function semantics with only a common selection function for the two operators? The answer is positive, and in the rest of this section we show how to modify the conditions on the selection function f to recover completeness. All we have to do is to replace the condition E3 with the following condition:

$$F1 \quad \text{If } \|A\| \neq \emptyset, \text{ then, for all } w, f(\|A\|, w) \subseteq \|A\|; \text{ otherwise } w \notin f(\|A\|, w).$$

It is immediate to give a counter-model for $\neg C\perp$: Let $W = \{w_1, w_2\}$ and $f(\emptyset, w_1) = \{w_2\}$. Since $f(\emptyset, w_1) \neq \emptyset$, and $w_1 \notin f(\emptyset, w_1)$ we have that $w_1 \models C\perp$.

As a first result for this semantics we show that axioms are valid in it and the inference rules preserve validity. We use \mathcal{S} to denote an Elgesem model that satisfies condition F1.

THEOREM 7. *If $\vdash_{\mathcal{L}_1} A$ then $\models_{\mathcal{S}} A$.*

⁶The condition that $\|A\|^\mathcal{E} \neq W$ is due to the axiom A1, which requires it.

Proof. The only not trivial case is that of axiom A2, since its characteristic condition E2 and F1 are entangle together in this semantics. Condition E2 takes care of the majority of cases, but we have to be careful since it is possible that the conjunction of A and B is inconsistent. If $w \models EA \wedge EB$, then $w \in \llbracket EA \wedge EB \rrbracket$; thus $w \in \llbracket EA \rrbracket \cap \llbracket EB \rrbracket$, which means that $w \in f(\llbracket A \rrbracket, w)$ and $w \in f(\llbracket B \rrbracket, w)$. According to condition F1 we have $\llbracket A \rrbracket \neq \emptyset$ and $\llbracket B \rrbracket \neq \emptyset$, which implies that $f(\llbracket A \rrbracket, w) \subseteq \llbracket A \rrbracket$ and $f(\llbracket B \rrbracket, w) \subseteq \llbracket B \rrbracket$. On the other hand it is possible that $\llbracket A \wedge B \rrbracket = \emptyset$, which means that $w \notin f(\llbracket A \wedge B \rrbracket, w)$. If this is the case then $\llbracket A \rrbracket \cap \llbracket B \rrbracket = \emptyset$; hence $f(\llbracket A \rrbracket, w) \cap f(\llbracket B \rrbracket, w) = \emptyset$. On the other hand if $\llbracket A \rrbracket = \emptyset$ (or $\llbracket B \rrbracket = \emptyset$) then $\llbracket A \wedge B \rrbracket = \emptyset$ and so $f(\llbracket A \rrbracket, w) = f(\llbracket A \wedge B \rrbracket, w)$. $A \equiv B$ iff $\llbracket A \rrbracket = \llbracket B \rrbracket$. In particular if $\llbracket A \rrbracket = \llbracket B \rrbracket = \emptyset$, then $f(\llbracket A \rrbracket, w) = f(\llbracket B \rrbracket, w)$. \square

The proof for the completeness is based on canonical models.

DEFINITION 8. A *selection function canonical model* is a structure $\mathcal{S}_c = \langle W, f, v \rangle$ such that:

- W is the set of all \mathcal{L}_1 -maximal consistent sets;
- v is an Elgesem valuation function such that, for all atomic proposition p , $w \models p$ iff $p \in w$.
- $f : \mathcal{P}(W) \times W \mapsto \mathcal{P}(W)$ is thus defined:
 - if $CA \notin w$, then $f([A]^{\mathcal{S}_c}, w) = \emptyset$; otherwise
 - if $[A]^{\mathcal{S}_c} = \emptyset$, then $f([A]^{\mathcal{S}_c}, w) = W - \{w\}$,
 - if $[A]^{\mathcal{S}_c} \neq \emptyset$, then $f([A]^{\mathcal{S}_c}, w) = [EA]^{\mathcal{S}_c}$.

where $[A]^{\mathcal{S}_c}$, the membership set of a formula A , is defined as follows: $[A]^{\mathcal{S}_c} = \{w \in W : A \in w\}$.

An immediate consequence of the above construction and Lindebaum's Lemma is the following proposition.

PROPOSITION 9. Let \mathcal{S}_c be a canonical selection function model $\langle W, f, v \rangle$, then:

- $[A]^{\mathcal{S}_c} = \emptyset$ iff $A \equiv \perp$.
- $|W| > 1$.
- If $A \not\equiv \top$ and $A \not\equiv \perp$, then $[EA]^{\mathcal{S}_c} \neq \emptyset$.

LEMMA 10. For every world $w \in W$ in \mathcal{S}_c , and every formula A , $w \models_{\mathcal{S}_c} A$ iff $A \in w$.

Proof. We prove the lemma by induction on the complexity of the formula. The inductive base is given by the basic condition on the valuation function for canonical models. Furthermore we consider only the case of modal operators.

If $w \models EA$, then by the evaluation function we have $w \in f(\llbracket A \rrbracket, w)$; by the inductive hypothesis $w \in f(\llbracket A \rrbracket, w)$, thus $w \in [EA]$, therefore $EA \in w$.

If $EA \in w$, then this implies that $CA \in w$ and $A \in w$. Since w is consistent $A \not\equiv \perp$ and $[A] \neq \emptyset$; thus $f(\llbracket A \rrbracket, w) = [EA]$ and then $w \in f(\llbracket A \rrbracket, w)$. By the inductive hypothesis $w \in f(\llbracket A \rrbracket, w)$, which implies $w \models EA$.

If $w \models CA$ then $f(\|A\|, w) \neq \emptyset$, and by the inductive hypothesis $f([A], w) \neq \emptyset$; by construction $CA \in w$.

If $CA \in w$, then either $f([A], w) = [EA]$ or $f([A], w) = W - \{w\}$. Clearly A cannot be \top , thus, according to Proposition 9, $f([A], w) \neq \emptyset$, and by the inductive hypothesis so is $f(\|A\|, w)$; therefore $w \models CA$. \square

LEMMA 11. \mathcal{S}_c satisfies conditions E1, E2, and F1.

Proof. $\neg C\top$ is an axiom, so $\neg C\top \in w$, for every world w ; hence $C\top \notin w$. By the construction of canonical models we have $f([\top], w) = \emptyset$. Since $[\top] = W$, we have $f(W, w) = \emptyset$.

If $w \in f([A], w) \cap f([B], w)$, then $w \in f([A], w)$ and $w \in f([B], w)$. This means that $[A] \neq \emptyset$ and $[B] \neq \emptyset$. From this we obtain that $EA \in w$ and $EB \in w$. Consequently $EA \wedge EB \in w$ and by the property of maximal consistent sets $E(A \wedge B) \in w$. All we have to prove now is that $[A \wedge B] \neq \emptyset$. To prove it we can use the same argument we have developed in the proof of Theorem 7 when we have shown that $EA \wedge EB \rightarrow E(A \wedge B)$ is valid.

If $A \equiv \perp$ then either $f([A], w) = W - \{w\}$ or $f([A], w) = \emptyset$. In both cases $w \notin f([A], w)$. If $A \not\equiv \perp$, then, if $CA \in w$, $f([A], w) = [EA]$. But for every world x if $EA \in x$ then $A \in x$; therefore $f([A], w) \subseteq [A]$. On the other hand if $CA \notin w$, then $f([A], w) = \emptyset$, thus, trivially $f([A], w) \subseteq [A]$. \square

From the two Lemmata above we obtain that \mathcal{L}_1 is complete with respect to \mathcal{S} .

THEOREM 12. $\vdash_{\mathcal{L}_1} A$ iff $\models_{\mathcal{S}} A$.

5 Non-normal Worlds and Relational Models

In the previous sections we have examined Elgesem's modal logic of agency and ability using semantics with different flavours. In general the selection function semantics and neighbourhood semantics give raise to the same structure: the selection function semantics focuses on the worlds where some actions can be realized in relation to a given world, while the neighbourhood semantics identifies the actions (formulas) that can be completed successfully in a given world.

In Section 4 we proposed a characterization of \mathcal{L}_1 based on models satisfying condition F1. According to the intended reading $f(\emptyset, w)$ is the set of worlds where the agent realizes his/her ability to bring about an impossible goal (whatever an impossible goal is). So in some senses, $f(\emptyset, w)$ corresponds to a set of impossible or imaginary worlds⁷. At any rate, the technical machinery of impossible (non-normal, queer) worlds offers us the opportunity to present an alternative class of Elgesem's models for \mathcal{L}_1 . All we have to do is to supplement the set W of possible worlds with the impossible world w_\perp , to establish that for every formula A , $w_\perp \models A$, and to define validity as validity at the normal worlds. The revised semantics makes explicit the need for impossible worlds –after all, if we assume that agents might have the ability to realize the impossible, it seems plausible to have a semantic counterpart for this notion. Hansson and Gärdenfors [10] point out that it is possible to destroy the general dependency of modal operators on the underlying semantic structure (in the case at hand the selection function f , and

⁷It is beyond the scope of the paper to give a characterisation of impossible worlds. All we ask for is that non-normal/impossible worlds are worlds whose rules and laws are different from the rules and laws of the normal worlds.

the accessibility relation R in relational models) by using non-normal/impossible worlds obeying to different logical rules.

Technically non-normal worlds deny the general idea behind intensional semantics that the value of modal formulas at a world w depends on the values of other formulas in other worlds, and validity is defined as validity at the normal worlds. Although the philosophical intuition behind non-normal worlds is sound, it commits us to postulate their existence; what is more is that its treatment is rather unsatisfactory: they are taken as black-boxes without any further analysis of their (internal) structure. In this way, we fail to recognise the potential multiplicity of types of non-normal worlds. A more appropriate solution is to recast the semantics with some more general type of dependence relation between truth of modal formulas and truth in other worlds [10].

Scott-Montague models have been devised, originally, to overcome the drawback of non-normal worlds we just have alluded to; but, for Elgesem's models, we have to reintroduce them, either implicitly or explicitly. If we have to reinstate non-normal/impossible worlds in order to prevent $\neg C \perp$ to be valid in Elgesem's models, then we overstep the very own idea motivating this type of semantics.

Since non-normal worlds are required, either implicitly or explicitly, in Elgesem's models the advantages of using a selection function semantics instead of relational models with non-normal worlds is lost. One could then ask if it is possible to devise a relational model for \mathcal{L}_1 (and \mathcal{L}_2). In the rest of this section we will investigate this issue.

Classical modal logics are characterised by models with the following structure [8]: $\langle W, N, R^*, \nu \rangle$ where W, ν are as before, $N \subseteq W$ is the set of normal worlds, and R^* is a set of binary relations over $N \times W$. The valuation clause for \Box is

$$w \models \Box A \text{ iff } w \in N \text{ and } \exists R \in R^* \text{ such that } \forall x (wRx \text{ iff } x \models A) \quad (8)$$

The set of non-normal world is denoted by Q (where $Q = W - N$). Alternatively we could define a model as $\langle W, Q, R^*, \nu \rangle$. Clearly if $w \in Q$, for any formula A , $w \not\models \Box A$. Worlds in Q corresponds to worlds in a neighbourhood model with empty neighbourhoods.

Now to accommodate C and E we have to combine one model for the E component and one model for the C components. Fortunately the two operators are related by axiom A4, thus we can adopt the structure⁸ $\langle W, Q^E, Q^C, R^E, R^C, \nu \rangle$ where W is a set of possible worlds, Q^E and Q^C are set of non-normal worlds such that $Q^C \subseteq Q^E$, R^E and R^C are sets of binary relations with signature $W - Q^X \times W$, and ν is an assignment. Moreover

- R1 $\forall R \in R^C \forall w \exists x \neg (wRx)$ (all relations in R^C are point-wise non-universal);
- R2 $\forall w \notin Q^E \forall R, S \in R^E \exists T \in R^E$ such that $R_w \cap S_w = T_w$ (R^E is point-wise closed under intersection);
- R3 $\forall R \in R^E \forall w (wRw)$ (all relations in R^E are reflexive);
- R4 $\forall w \notin Q^E \forall R \in R^E \exists R' \in R^C$ such that $R_w = R'_w$ (the relations in R^E are sub-relations of relations in R^C);

⁸From now on we will use X as a variable ranging over C, E .

R5 $\forall R \in R^C \forall w \exists x (wRx)$ (all relations in R^C are serial).

As we shall see \mathcal{L}_1 is determined by the class of relational models satisfying R1–R4, while \mathcal{L}_2 by R1–R5. To prove these results we are going to show that for each relational model there is an equivalent neighbourhood model, and for every (finite) neighbourhood model there is an equivalent relational model.

Before proving this result we give an auxiliary lemma about sufficient conditions to ensure the equivalence of relational and neighbourhood models. In what follows we will use R_w , for $R \in R^X$ to denote the set of worlds accessible from w using the relation R , formally: if $R \in R^X$, then $R_w = \{w' \in W : wRw'\}$.

LEMMA 13. *Let $\mathcal{N} = \langle W, N^E, N^C, \nu \rangle$ be a neighbourhood model and $\mathcal{R} = \langle W, Q^E, Q^C, R^E, R^C, \nu \rangle$ be a relational model such that*

1. $\forall w \in W$ if N_w^X , then $\forall x \in N_w^X \exists R \in R^X$ such that $x = R_w$, and
2. $\forall w \in W$ if $w \notin Q$, then $\forall R \in R^X \exists x \in N_w^X$ such that $x = R_w$.

Then for all formulas A : $\models_{\mathcal{N}} A$ iff $\models_{\mathcal{R}} A$.

Proof. The proof is by induction on the complexity of A . The two models have the same set of possible worlds and the same assignment, thus they agree on every propositional variable. For the inductive step and the modal operators all we have to do is to apply the condition 1 and 2. \square

For every relational model we can generate an equivalent neighbourhood model where $N_w^X = \{R_w : R \in R^X\}$. For the other direction, on the other hand, we have to be careful. Beside the constraints dictated by the internal structure of the model we have to ensure that the set of relations generated from N_w^E is closed under intersection and the relations are serial if we want to satisfy R5. The idea is the same as in the other direction: we use the sets in N_w^X to create instances of relations in R^X . Here the problem is that given two worlds w and w' it is very likely that $|N_w^E| \neq |N_{w'}^E|$; hence w generates $|N_w^E|$ sub-relations and w' generates $|N_{w'}^E|$ sub-relations, thus there are sub-relations without elements in relation with w . A simple solution to obviate this problem is to pick a fixed but arbitrary $x \in N_w^E$ for all the additional relations.

THEOREM 14.

1. For every (finite) relational model \mathcal{M} there is an equivalent (finite) neighbourhood model \mathcal{N} such that if \mathcal{R} satisfies Rn then \mathcal{N} satisfies Cn (for $1 \leq n \leq 5$).
2. For every finite neighbourhood model \mathcal{N} there is an equivalent finite relational model \mathcal{R} such that if \mathcal{N} satisfies Cn then \mathcal{R} satisfies Rn (for $1 \leq n \leq 5$).

Proof. First of all the models will have the same set of worlds and the same assignment, thus all we have to show is that it is possible to generate appropriate sets of relations from the given neighbourhood functions and appropriate neighbourhood functions from the given sets of relations.

Part 1. Given a (finite) relational model \mathcal{R} we can generate an equivalent (finite) neighbourhood model as follows:

- If $w \in Q^X$ then $N_w^X = \emptyset$; otherwise
- $N_w^X = \{R_w : R \in R^X\}$.

It is immediate to verify that the conditions of Lemma 13 are satisfied by the models obtained from the above construction, therefore the generated models are equivalent to the generating models.

Part 2. To build a finite relation model from a finite neighbourhood model we use the following construction.

For each N_w^E and N_w^C let Σ_w^E and Σ_w^C be sequences of all the elements in N_w^E and N_w^C such that if $i \leq |N_w^E|$, then $\Sigma_{w,i}^E = \Sigma_{w,i}^C$ (we use $\Sigma_{w,i}^X$ to indicate the i -th element of Σ_w^X). Moreover

$$e = \max \{ |N_w^E| : w \in W \} \quad c = \max \{ |N_w^C| : w \in W \}.$$

Then

$$R^E = \bigcup_{1 \leq i \leq e} R_i^E \quad R^C = \bigcup_{1 \leq i \leq c} R_i^C$$

where

$$R_i^E = \{ (w, w') : w \notin Q^E \text{ and } w' \in \alpha(w, i) \}$$

$$R_i^C = \{ (w, w') : w \notin Q^C \text{ and } w' \in \gamma(w, i) \}$$

where α and γ are partial functions with signature $\alpha : W \times \mathbb{N} \mapsto N^E$ and $\gamma : W \times \mathbb{N} \mapsto N^C$ such that:

$$\alpha(w, i) = \begin{cases} \text{undefined} & \text{if } i > e \text{ or } N_w^E = \emptyset \\ \Sigma_{w,i}^E & \text{if } i \leq |N_w^E| \\ \Sigma_{w,1}^E & \text{otherwise} \end{cases}$$

and

$$\gamma(w, i) = \begin{cases} \text{undefined} & \text{if } i > e + c \text{ or } N_w^C = \emptyset \\ \Sigma_{w,i}^C & \text{if } i \leq |N_w^E| \\ \Sigma_{w,i-e+|N_w^C|}^C & \text{if } e < i \leq e + |N_w^C| - |N_w^E| \\ \Sigma_{w,1}^C & \text{otherwise} \end{cases}$$

It is easy to verify that the models obtained from the above construction obey to the conditions of Lemma 16; consequently this construction produces equivalent models. \square

Due to the above procedure to generate such relational models, in the case of infinite \mathcal{N} or \mathcal{N}' models we would get non-enumerable infinitary relational structures. To avoid these complexities, it is sufficient to consider \mathcal{N} and \mathcal{N}' when they are finite. This is possible by preliminarily showing that \mathcal{L}_1 and \mathcal{L}_2 have the finite model property wrt the neighbourhood models previously defined. The fmp follows immediately from the results of Lewis [16] and [25] that every classical non-iterative modal logic has the finite model property.⁹ Clearly \mathcal{L}_1 and \mathcal{L}_2 are non-iterative thus we have the following theorem.

⁹A modal logic is non-iterative iff it can be axiomatized by using only non-iterative axioms. A formula (axiom) A is non-iterative iff for every subformula $\Box_i B / \Diamond_i B$ of A , B does not contain a modal operator.

THEOREM 15. \mathcal{L}_1 and \mathcal{L}_2 have the fmp.

We can now prove the completeness of the \mathcal{L}_1 and \mathcal{L}_2 with respect to the relational models developed in this section.

THEOREM 16. Let \mathcal{R}_1 be a relational model satisfying R1–R4, and \mathcal{R}_2 be a relational model satisfying R1–R5; then

1. $\vdash_{\mathcal{L}_1} A$ iff $\models_{\mathcal{R}_1} A$;
2. $\vdash_{\mathcal{L}_2} A$ iff $\models_{\mathcal{R}_2} A$.

Proof. Let us consider only \mathcal{L}_1 . From Theorem 3 we know that $\models_{\mathcal{N}} A \rightarrow \vdash_{\mathcal{L}_1} A$, which is equivalent to saying that $\not\vdash_{\mathcal{L}_1} A \rightarrow \not\models_{\mathcal{N}} A$. Since \mathcal{L}_1 has the finite model property, there is a finite model \mathcal{N}_{FIN} and a world w in it such that $w \models_{\mathcal{N}_{FIN}} \neg A$. According to Proposition 14 and the generation of the corresponding relational model $w \models_{\mathcal{R}_1} \neg A$ which implies $\not\vdash_{\mathcal{R}_1} A$. Then, $\not\vdash_{\mathcal{L}_1} A \rightarrow \not\vdash_{\mathcal{R}_1} A$ and so $\models_{\mathcal{R}_1} A \iff \vdash_{\mathcal{L}_1} A$. The proof for \mathcal{L}_2 and \mathcal{R}_2 is analogous. \square

Here we want to propose a simple interpretation of relational models: the capability of an agent to realize a particular state A depends on his/her ability to perform some actions in the situation described by the then actual world. Accordingly each accessibility relation corresponds to a concrete action. In this perspective non-normal worlds are just situations where an agent has no possibility to perform any action.

6 Discussion

When we consider the semantics developed by Elgesem we have to notice that he uses only one selection function to represent the two modal operators instead of the two neighbourhood functions of Section 3. This amounts to say that Elgesem considers agency and ability as two facets of the same phenomenon –the phenomenon described by the selection function. Thus to discern the two concepts he has to adopt two different valuation clauses. In particular the condition for E is the condition for a \square operator, while that for C is the condition used for a \diamond operator. However these conditions, in the context of non-normal modal logic, do not imply that \diamond is the dual of \square . On the contrary the neighbourhood semantics assumes two separate but related modal operators.

It has been argued that an agent can carry out an action successfully if she has the ability as well as the opportunity to do it. Indeed Elgesem studies the relationships between ability and agency, and he correctly realizes that agency implies opportunity, i.e., $EA \rightarrow OpA$, where Op is the modal operator for opportunity. But the notion of opportunity is given in terms of agency, i.e., $OpA \equiv (E\neg A \vee A)$. Therefore we believe that the semantics proposed by Elgesem does not fully capture the idea that agency consists of ability plus opportunity since those three notions are represented by the same selection function. The other semantics do recognise that ability alone is not enough to represent agency and that it has to be supplemented by something else.

Finally Elgesem semantics requires the introduction (either implicitly or explicitly) of non-normal worlds, but their interpretation is not satisfactory; on the contrary the interpretation we have proposed for non-normal worlds seems to fit nicely with the intended reading of the accessibility relations for this type of logics.

References

- [1] Nuel Belnap and Michael Perloff. Seeing to it that: A canonical form for agentives. *Theoria*, 54:175–99, 1988.
- [2] Nuel Belnap and Michael Perloff. The way of the agent. *Studia Logica*, 51:463–484, 1992.
- [3] José Carmo and Olga Pacheco. Deontic and action logics for organized collective agency modeled through institutionalized agents and roles. *Fundamenta Informaticae*, 48:129–163, 2001.
- [4] Brian Chellas. *The Logical Form of Imperatives*. Perry Lane Press, Palo Alto, 1969.
- [5] Brian Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.
- [6] Dag Elgesem. *Action Theory and Modal Logic*. Phd, Institut for filosofi, Det historisk-filosofiske fakultetet, Universitetet i Oslo, 1993.
- [7] Dag Elgesem. The modal logic of agency. *Nordic Journal of Philosophical Logic*, 2(2):1–46, 1997.
- [8] Olivier Gasquet and Andreas Herzig. From classical to normal modal logic. In Heinrich Wansing, editor, *Proof Theory of Modal Logic*, pages 293–311. Kluwer, Dordrecht, 1996.
- [9] Jonathan Gelati, Guido Governatori, Antonino Rotolo, and Giovanni Sartor. Declarative power, representation, and mandate: A formal analysis. In Trevor Bench-Capon, Aspasia Daskalopulu, and Radboudb Winkels, editors, *Legal Knowledge and Information Systems*, pages 41–52. IOS Press, Amsterdam, 2002.
- [10] Bengt Hansson and Peter Gärdenfors. A guide to intensional semantics. In *Modality, Morality and Other Problems of Sense and Nonsense. Essays Dedicated to Sören Halldén*, pages 151–167. Gleerup, Lund, 1973.
- [11] Risto Hilpinen. On action and agency. In E. Ejerhed and S. Lindström, editors, *Logic, Action and Cognition: Essays in Philosophical Logic*, pages 3–27. Kluwer Academic Publishers, Dordrecht, 1997.
- [12] Jaakko Hintikka. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4:475–484, 1975.
- [13] John F. Horty and Nuel Belnap. The deliberative stit. a study of action, omission, ability, and obligation. *Journal of Philosophical Logic*, 24:583–644, 1995.
- [14] Andrew J. I. Jones and Marek Sergot. A formal characterisation of institutionalised power. *Journal of IGPL*, 3:427–443, 1996.
- [15] Andrew J.I. Jones. A logical framework. In Jeremy Pitt, editor, *Open Agent Societies: Normative Specifications in Multi-Agent Systems*, chapter 3. John Wiley and Sons, Chichester, 2003.
- [16] David Lewis. Intensional logic without iterative axioms. *Journal of Philosophical Logic*, 3(4):457–466, 1974.
- [17] Ingmar Pörn. *The Logic of Power*. Blackwell, Oxford, 1970.
- [18] Ingmar Pörn. *Action Theory and Social Science: Some Formal Models*. Reidel, Dordrecht, 1977.
- [19] Filipe Santos and José Carmo. Indirect action. influence and responsibility. In Mark Brown and José Carmo, editors, *Deontic Logic, Agency and Normative Systems*. Springer, Berlin, 1996.
- [20] Filipe Santos, Andrew J.I. Jones, and José Carmo. Action concepts for describing organised interaction. In *Thirtieth Annual Hawaii International Conference on System Sciences*. IEEE Computer Society Press, Los Alamitos, 1997.
- [21] Krister Segerberg. Bringing it about. *Journal of Philosophical Logic*, 18:327–347, 1989.
- [22] Krister Segerberg. Getting started: Beginnings in the logic of action. *Studia Logica*, 51:347–358, 1992.
- [23] Marek Sergot. A computational theory of normative positions. *ACM Transactions on Computational Logic*, 2(581–622), 2001.
- [24] Marek Sergot and Fiona Richards. On the representation of action and agency in the theory of

- normative positions. *Fundamenta Informaticae*, 48:273–293, 2001.
- [25] Timothy J. Surendonk. Canonicity for intensional logics without iterative axioms. *Journal of Philosophical Logic*, 26(4):391–409, 1997.