Journal of Philosophical Logic (2005) 34: 403–431
DOI: 10.1007/s10992-004-6368-1

GUIDO GOVERNATORI and ANTONINO ROTOLO

ON THE AXIOMATISATION OF ELGESEM'S LOGIC OF AGENCY AND ABILITY

ABSTRACT. In this paper we show that the Hilbert system of agency and ability presented by Dag Elgesem is incomplete with respect to the intended semantics. We argue that completeness result may be easily regained. Finally, we shortly discuss some issues related to the philosophical intuition behind his approach. This is done by examining Elgesem's modal logic of agency and ability using semantics with different flavours.

1. INTRODUCTION

Modal logic of agency is a traditional research field in philosophical logic (for a modern history of this field see [32]). 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 [33], the main limit of this approach, as found in the literature, is that it is "too abstract". For example, it does not usually capture the difference between the modal qualifications "sees to it" and "brings it about". Both expressions are in general represented by modal operators with the same logical properties, despite the fact that the former exhibits a clear intentional character, whereas the latter may refer as well to unintentional actions [15]; thus the addition of a modal operator for intention is required to disambiguate the two readings [13].

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 [18, 19, 28, 29]. 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 [33]. As 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., [5, 11, 12, 18, 19, 28, 29]).

The formal properties of modal logic of agency have been extensively investigated, and a number of variants and axiomatisations can be found in literature (see, e.g., [2, 3, 6, 8, 9, 17, 26–29, 31, 32]). Despite this great variety, it is possible to identify a minimal core of axioms that seem to characterise indisputably some aspects of agency. The recent contributions by Dag Elgesem are meant to work in this direction [8, 9].

The paper is organised as follows: in Section 2 we rehearse the basic notions for a modal logic of agency and ability; then in Section 3 we present the account given by Elgesem: in particular we will introduce the class of selection function models proposed by Elgesem to explain, semantically, the notions of agency and ability based on a goal directed interpretation of such concepts; we also discuss the corresponding axiomatisation. As we will see, the semantics validates the formula $\neg C\bot$, whose interpretation is that no agent has the ability to realise the impossible. However, in Section 5 we show that $\neg C\bot$ cannot be derived from the axiom system presented in Section 3. To this end, we introduce neighbourhood semantics, we prove that there is a class of neighbourhood models characterising the axiom system, and we build a model falsifying $\neg C\bot$. As a consequence, the proposed axiomatisation is incomplete with respect to the intended semantics. Moreover we give conditions to transform a neighbourhood model into an equivalent selection function model and the other way around. It turns out that we can restore completeness by adding $\neg C\bot$ as an additional axiom. This leaves us with two logics: one with $\neg C\bot$ and one without it. Accordingly, we discuss, in Section 4, some philosophical issues related to the interpretation of the notion of ability when one accepts or rejects the above formula. Since both logics are intuitively acceptable given the proper interpretations, in Section 6 we investigate whether it is possible to regain completeness for the logic without $\neg C\bot$ using the type of models proposed by Elgesem. To this end we have to introduce, in an implicit manner, non-normal possible worlds (i.e., worlds with "special" conditions to evaluate formulas). As we will argue this is against the very own idea of the neighbourhood and selection function semantics; thus in Section 7 we develop a new type of relational (Kripke-style) possible world semantics with non-normal worlds, and we prove soundness and completeness for the two logics of agency and ability with respect to this semantics. We conclude the paper (Section 8) with a brief discussion about the choice of semantics for modal logics and we give some hints for further philosophical and technical investigations.

The focal point of the paper revolves around the incompleteness of the logic proposed by Elgesem with respect to the intended semantics. This has interesting ramifications on both philosophical and technical issues. Philosophically our analysis may shed light on different meanings of agency while, technically, it opens questions on the intuitive appeal of different types of possible world semantics.

2. Modal Logic of Agency and Ability

We will focus here on two praxeological notions among those considered by Elgesem.[1] 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 $E_i A$ means that the agent i brings it about that A. Elgesem's logic of E is a classical non-normal system [7], namely is closed under logical equivalence, and is characterised by the following schemas.

(1) $E_i A \rightarrow A$

(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 that is behind the common reading of "bring about" concept.

(2) $\neg E_i \top$

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 (causal) 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. This means that an agent i can only realise something which is potentially avoidable. In this perspective, no agent can bring about what is logically unavoidable. Accordingly $\top$ cannot be realised with any contribution of i because its occurrence is (causally) independent of any action of i.

(3) $(E_i A \wedge E_i B) \rightarrow E_i(A \wedge B)$

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 [8, 9].

Notice that this minimal core of principles has been recognised also by [28, 29]. The main difference between them and Elgesem regards the characterisation of the interplay and influence between more agents. In [28, 29] it is accepted

(4) $E_i E_j A \rightarrow \neg E_i A$

to emphasise that a formula like $E_i A$ strongly expresses the idea that the agent i brings it about that A *directly* and *personally*: If i makes so that j brings it about that A, then it is not possible to say that i realises A, since such a state of affairs is achieved directly by j. More weakly, but in a similar perspective, Elgesem simply rejects the schema

(5) $E_i E_j A \rightarrow E_i A$

which is adopted, for example, by Chellas [6].

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, $C_i A$ 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.

(6) $E_i A \rightarrow C_i A$

This schema states a strong connection between ability and agency: If i realises successfully A, it is obvious that i is able to do this.

(7) $\neg C_i \top$

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 world.[2]

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.

3. An Axiomatisation for Agency and Ability

Elgesem's analysis starts from semantical considerations [8, 9]. 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 $\mathscr{E}$ is a structure

(8) $\langle W, f, v\rangle$

where W is a (non-empty) set of possible worlds, f is a selection function with signature $\mathscr{P}(W) \times W \mapsto \mathscr{P}(W)$, 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 realised.

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. Let $\mathscr{M}$ be a model and A be a formula. The truth set of A wrt to $\mathscr{M}$, $\|A\|^{\mathscr{M}}$ is thus defined:

$$\|A\|^{\mathscr{M}} = \{w \in W : w \vDash_{\mathscr{M}} A\}.$$

An Elgesem model is a selection function model $\mathscr{E}$ satisfying the following valuation clauses (from now on, whenever clear from the context we drop subscripts and superscripts):

S1. $w \vDash_{\mathscr{E}} p$ iff $w \in v(p)$;
S2. $w \vDash_{\mathscr{E}} \neg A$ iff $w \nvDash_{\mathscr{E}} A$;
S3. $w \vDash_{\mathscr{E}} A \rightarrow B$ iff $w \nvDash_{\mathscr{E}} A$ or $w \vDash_{\mathscr{E}} B$;
S4. $w \vDash_{\mathscr{E}} EA$ iff $w \in f(\|A\|^{\mathscr{E}}, w)$;

S5. $w \models_{\mathscr{E}} CA$ iff $f(\|A\|^{\mathscr{E}}, w) \neq 0$.

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 (6), 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 realises 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 realised in every world is not a state the agent is able to bring about. As an immediate consequence of this constraint we have the validity of (7) 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 realises his/her ability to bring about a goal the goal is indeed realised. It is easy to see that it validates the success principle (1).

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

A0 propositional logic,
A1 $\neg C\top$,
A2 $EA \wedge EB \rightarrow E(A \wedge B)$,
A3 $EA \rightarrow A$,
A4 $EA \rightarrow CA$;

plus Modus Ponens and

$$(9) \qquad \frac{A \equiv B}{EA \equiv EB} RE_E \qquad \frac{A \equiv B}{CA \equiv CB} RE_C$$

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

$$(10) \qquad \neg E\bot$$

This principle states that nobody can realise an inconsistent (impossible) state. But, what about the corresponding principle that nobody is capable to produce an inconsistent state?

$$(11) \qquad \neg C\bot$$

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

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

$$\begin{aligned} w \models_{\mathcal{E}} \neg C\bot &\iff w \not\models_{\mathcal{E}} C\bot \\ &\iff f(\|\bot\|^{\mathcal{E}}, w) = \emptyset. \end{aligned}$$

According to condition E3

$$\forall w \in W, \quad f(X, w) \subseteq X$$

and, $\|\bot\|^{\mathcal{E}} = \emptyset$; hence

$$f(\|\bot\|^{\mathcal{E}}, w) \subseteq (\|\bot\|^{\mathcal{E}}) = \emptyset.$$

According to the intended interpretation, $\neg C\bot$ means that an agent is not able to realise 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 [16] 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.

In the next section we will provide some brief philosophical comments on whether adopting or rejecting schema (11).

4. Impossible Ability?

The interpretation assigned by Elgesem to the "bringing about" operator is that the agent realises a state of affair A by practically or causally contributing to its occurrence [9, pp. 19ff.]. More precisely, Elgesem says that agent's actions be causally necessary conditions for the goal-event A to occur. Agent's actions are of course necessary but not sufficient because the effective realisation of A depends also on some environmental factors, namely, on external circumstances that allow for the exercise of agent's ability to achieve A. Within this specific background, and taking into account that we are dealing with successful actions, it seems obvious that the notion of avoidability should characterise agency because it does not make sense that an agent practically contributes to the occurrence of $\top$, this last being unavoidable: the occurrence of $\top$ is entirely independent of any action of the agent. For similar – and perhaps stronger – reasons, $\neg E\bot$, which is a theorem in Elgesem's axiomatisation as is trivially derived from $EA \rightarrow A$, is reasonable because it is an absurdity that an agent successfully realises the impossible, which, by definition, cannot be practically or causally realised.

What about the notion of practical ability? The idea of avoidability should also be applied, as Elgesem does, to the operator C: the occurrence of a state of affairs A cannot be caused by one agent if the goal obtains in every state of the world. But a similar rationale should be adopted also when the goal is $\bot$ and, in fact, we will show that (11) is technically required in Elgesem's axiomatisation.[4] But let us forget for a while that Elgesem's logic is incomplete without (11). Hence, the question is: In which sense does an agent have the ability to cause the occurrence of the impossible? Here the point may be more subtle than it appears. As we know, in virtue of (6), agency implies ability since any action of an agent realising A is successful and the occurrence of A depends on her action: if I realise A this requires that I am able to do it. This is obvious and, in fact, given $E\bot \rightarrow C\bot$, we cannot infer by detachment $C\bot$ since $E\bot$ is equivalent to $\bot$. However, if all performed actions require a corresponding ability, this does not require that all abilities are exercised: the domain of abilities just includes that of actions. We may argue thus that there is a state of affairs, say A, such that some agents are able to realise it, but that cannot be effectively realised. What does this mean? Bear in mind that the idea that an agent can effectively realise a state of affairs may be obtained as the combination of ability plus opportunity: an agent having the ability to do A may be prevented by circumstances from exercising this ability (see also [20]). On this interpretation, the ability to do $\bot$ can be viewed as an ability such that, for all kind of circumstances, these last prevent any agent to exercise such an ability. Notice that, in Elgesem's analysis, the opportunity to do something excludes that this can be $\bot$. In fact in Elgesem's logic [9, pp. 33 and 35] we have

(12) $\neg \text{Opportunity}_i E_i \bot$

which is equivalent to

(13) $\neg \text{Opportunity}_i \bot$

In other words no agent "can effectively" do the impossible.

But the question, again, is: Does it make sense to maintain that a state of affairs A cannot be practically realised because all possible circumstances prevent the exercise of the ability to do A? The impression is that such an ability is void: that all possible circumstances prevent to realise A means that there is no circumstance that allows any agent to realise A (cf. [4, p. 18]). Therefore, the idea of having such an ability is meaningless and so the schema (11) is required.

In a more general perspective, we can add what Anthony Kenny [21, p. 214] writes about the impossibility to realise $\top$:

> The President of the United States has the power to destroy Moscow, i.e., to bring it about that Moscow is destroyed; but he does not have the power to bring it about that either Moscow is destroyed or Moscow is not destroyed. [...] [T]he power to bring it about that either p or not p is one which philosophers, with the exception of Descartes, have denied even to God.

As we tried to argue, this impossibility can be extended also to impossible states of affairs.

However, this is not the end of the story. Things may change when we assign to E and C a different meaning and we go beyond the idea of practical agency.

We may have a first exception when we deal with the idea of normative agency. In this perspective, the role of the schema $\neg C\bot$ can be debated if Elgesem's logic of agency is combined, as done in recent works on norm-governed agent systems, with deontic notions [5, 11, 12, 18, 19, 29]. A more extensive discussion of the interpretation of C would be to study it with regard to normative/deontic capability. We will confine ourselves to some short remarks. Indeed there are interpretations where $\neg C\bot$ is not an appropriate axiom. In many deontic logics the schema $\neg O\bot$ obtains and is meant to avoid the absurdity of norms that oblige to do something contradictory: these norms are useless since regulate something self-contradictory and because they cannot be accomplished with. In fact, it is quite common in the literature to assume (implicitly or explicitly) that the nature of norms is that they must be accomplishable [24].

In this context we may start from assuming that expressions such as COA and EOA mean respectively that an agent can draft a norm OA, and that the agent effectively issues OA (remember that E is a successful operator). However, since norms are the result of the exercise of a certain power, nothing in theory prevents a legislator to draft a norm like "it is obligatory to smoke and not to smoke".[5] It may be argued that the absurdity resides here in the obligation itself and not in the fact that such an obligation has been drafted. If this is reasonable, we may impose to have rational norms, and a consistent normative system, without having a rational legislator. This view makes explicit a possible way of distinguishing practical agency from normative agency: in fact, each of us, if legislator, can draft absurd norms whereas none of us is able to physically realise a state of affairs such as "I smoke and I do not smoke". If norms are required to be consistent (axiom D: $OA \rightarrow PA$ in case of normal modal deontic operator, or $\neg O\bot$), then $O\bot$ is equivalent to $\bot$ and so we have that $C\bot$ is consistent; hence (11) has to be rejected as a valid principle. This is indeed a possibility that we could admit, but, of course, we would presuppose a different conception of agency. If we move from C to E we will notice that $EO\bot$ is equivalent to $\bot$. This means that an agent (legislator) has the ability (power) to draft an inconsistent norm without making the normative system inconsistent. On the contrary, the legislator cannot issue, namely, make valid, an inconsistent norm without generating a contradiction within the normative system. To sum up, rejecting $\neg C\bot$ and adopting $\neg O\bot$ permits to distinguish between drafting norms and effectively issuing them within the normative system.

Other exceptions may be put forward when we try to use the notion of agency within more specific, and perhaps unexpected, contexts.

A somehow related interpretation arises in computer science where C, and E can be understood, respectively, as the permission (capability) to issue a computer instruction (let us say a syntactically correct line of code in a program), and to execute a computer instruction. In this reading $\bot$ corresponds to an illegal instruction, let us say a division by zero or an operation where two different values are assigned, at the same time, to one and the same memory register, which causes the system to crash, or a condition that violates some integrity constraints. In this interpretation $C\bot$ is clearly consistent since the mere fact that a potentially dangerous instruction has been inserted in some program does not imply that the instruction is actually executed by the interpreter or the compiler.

Similarly, in mathematics, C can be interpreted as an act of defining a property and E as an effective construction for the property. It is possible to provide a definition with an empty extension (thus corresponding to $\perp$, but then it is not possible to give any effective construction for such notion). For example, given the domain of rational numbers, any two rational numbers such that $x \neq y$, we define that x *divides* y iff $\exists z(x \cdot z = y)$. Hence $C(\mathit{divides}(0, 1))$ is consistent, since it corresponds to the definition just given, while $E(\mathit{divides}(0, 1))$ would be true if we can provide a number z such that $0 \cdot z = 1$; according to the axioms governing multiplication in this domain we have $\forall x(0 \cdot x = 0)$, from which we derive $\mathit{divides}(0, 1) = \perp$. From the axioms governing E we have $E\perp \rightarrow \perp$, and finally $\neg E\perp$, and so $E(\mathit{divides}(0, 1))$ is always false. We are aware that the parallelism between ability and agency, on the one hand, and defining and calculating in mathematics, on the other, may not convince the reader. In fact, such a comparison should be tested with regard to the full axiomatisation we could adopt in providing a precise support of this new reading of agency and ability. We will not enter here into further details. In our view, this last case can be roughly understood by applying the same intuition that is behind the previous example, which concerns computer science.

5. NEIGHBOURHOOD MODELS

As we have seen in the previous section $\neg C\perp$ is valid, but, as we will see, it is not provable from $\mathscr{L}_1$, hence $\mathscr{L}_1$ is incomplete wrt the intended semantics. To show that $\mathscr{L}_1$ is incomplete wrt $\mathscr{E}$ we have to provide a class of models such that $\mathscr{L}_1$ is complete for it and $\neg C\perp$ is false. While it is possible to devise a class of selection function models for $\mathscr{L}_1$ (see Section 6) we prefer to introduce models with a different structure. As 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. [14]).

A neighbourhood model $\mathscr{N}$ 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 $\mathscr{P}(\mathscr{P}(W))$, and v assigns subsets 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 $\mathscr{N} = \langle W, N^C, N^E, v\rangle$:

N1 $w \models_{\mathcal{N}} CA$ iff $\|A\|^{\mathcal{N}} \in N_w^C$;
N1 $w \models_{\mathcal{N}} EA$ 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. We provide the proof only for the correspondence between A4 and C4. For the other axioms and semantic conditions see [7, 30].

For the soundness part we have to show that axiom A4 is valid. Let us suppose it is not. Then there is a model satisfying conditions C1–C4, and that falsifies it; that is, there is a world w such that $w \not\models EA \rightarrow CA$. This means (1) $w \models EA$ and (2) $w \not\models CA$. From (1) we obtain $\|A\| \in N_w^E$, and then, by the inclusion condition C4, $\|A\| \in N_w^C$, but from (2) we have $\|A\| \notin N_w^C$. Hence we get a contradiction.

For the completeness part let us consider the minimal canonical neighbourhood model $\mathcal{N}_{min}$ for $\mathcal{L}_1$. $\mathcal{N}_{min}$ is defined as follows:

- W is the set of all $\mathcal{L}_1$-maximal consistent sets,
- $N_w^E = \{[A] : EA \in w\}$,
- $N_w^C = \{[A] : CA \in w\}$,
- $w \models p$ iff $p \in w$,

where $[A] = \{w \in W : A \in w\}$, and N^C and N^E satisfy conditions C1–C4.

Again we prove only the case for A4 and C4. Since the worlds in W are $\mathcal{L}_1$-maximal consistent sets, for any world w we have that either (1) $EA \in w$ or (2) $\neg EA \in w$.

For (1) we have

$$EA \in w \quad \text{iff} \quad [A] \in N_w^E$$

and, by the inclusion condition C4, $[A] \in N_w^C$; but

$$[A] \in N_w^C \quad \text{iff} \quad CA \in w.$$

Hence, both EA and CA are in w; therefore $EA \rightarrow CA \in w$.

On the other hand if (2) is the case, we can use the tautology $\neg EA \rightarrow (EA \rightarrow CA)$ to conclude that $EA \rightarrow CA \in w$.

In both cases $EA \rightarrow CA \in w$, and by the properties of canonical models we have $\vdash_{\mathcal{L}_1} EA \rightarrow CA$. □

It is easy to provide a neighbourhood model that falsifies $\neg C\bot$. Let $W = \{w\}$, $N^E_w = \emptyset$ and $N^C_w = \{\emptyset\}$. Here, $\|\bot\| = \emptyset \in N^C_w$, therefore $w \vDash C\bot$ and $w \nvDash \neg C\bot$. Hence we have the following result:

PROPOSITION 4. $\nvdash_{\mathscr{L}_1} \neg C\bot$.

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

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

(1) $w \in f(\|A\|^{\mathscr{E}}, w)$ *iff* $\|A\|^{\mathscr{N}'} \in N^E_w$; *and*
(2) $f(\|A\|^{\mathscr{E}}, w) \neq \emptyset$ *and* $\|A\|^{\mathscr{E}} \neq W$ *iff* $\|A\|^{\mathscr{N}'} \in N^C_w$.[6]

Then

$$\vDash_{\mathscr{E}} A \quad \textit{iff} \quad \vDash_{\mathscr{N}'} A.$$

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

Proof. Since the two models have the same worlds and the same assignment, clearly the two models agree on the valuation of propositional formulas. For the modal operators E and C we have

$$\begin{aligned} w \vDash_{\mathscr{E}} EA \quad &\text{iff} \quad w \in f(\|A\|^{\mathscr{E}}, w) \\ &\text{iff} \quad \|A\|^{\mathscr{N}'} \in N^E_w \\ &\text{iff} \quad w \vDash_{\mathscr{N}'} EA \end{aligned}$$

and

$$\begin{aligned} w \vDash_{\mathscr{E}} CA \quad &\text{iff} \quad f(\|A\|^{\mathscr{E}}, w) \neq \emptyset \text{ and } \|A\|^{\mathscr{E}} \neq W \\ &\text{iff} \quad \|A\|^{\mathscr{N}'} \in N^C_w \\ &\text{iff} \quad w \vDash_{\mathscr{N}'} CA \end{aligned}$$

For the other property we reason as follows: for every world w of $\mathscr{E}$ we have $f(\|\bot\|^{\mathscr{E}}, w) \subseteq \|\bot\|^{\mathscr{E}} = \emptyset$, so $f(\|\bot\|^{\mathscr{E}}, w) = \emptyset$, and from the relationships between the two models we obtain $\emptyset \notin N^C_w$. The other direction is immediate. □

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

C5 $\forall w, \emptyset \notin N^C_w$,

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

THEOREM 6.

(1) $\vdash_{\mathscr{L}_2} A$ *iff* $\vDash_{\mathscr{N}'} A$;
(2) $\vdash_{\mathscr{L}_2} A$ *iff* $\vDash_{\mathscr{E}} A$.

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

6. 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 [14], 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.[7] All we have to do is to replace condition E3 with the following condition:

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

It is immediate to give a counter-model for $\neg C\bot$: 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 $w_1 \vDash C\bot$.

As a first result for this semantics we show that axioms are valid in it and the inference rules preserve validity.

We use $\mathscr{S}$ to denote an Elgesem model that satisfies condition F1.

THEOREM 7. *If* $\vdash_{\mathscr{L}_1}$ *then* $\vDash_{\mathscr{S}} A$.

Proof. Clearly all propositional tautologies are valid and Modus Ponens preserves validity.

Axiom $EA \to A$. Let w be world in W. If $\|A\| \neq \emptyset$ then $f(\|A\|, w) \subseteq A$, and this is the well-known condition for this axiom to be valid, and thus true at w.

If $\|A\| = \emptyset$, then $w \notin f(\|A\|, w)$. But in this case $w \not\models EA$, and then $w \models EA \to A$.

Axiom $EA \to CA$. Let w be world in W. If $\|A\| \neq \emptyset$ then $f(\|A\|, w) \subseteq A$. If $\models EA$, then $w \in f(\|A\|, w)$; hence $f(\|A\|, w) \neq \emptyset$, $w \models CA$. Therefore $w \models EA \to CA$.

If $\|A\| = \emptyset$, then $w \notin f(\|A\|, w)$. But in this case $w \not\models EA$, and then $w \models EA \to CA$.

Axiom $\neg C\top$. This axiom is independent from the new condition.

Axiom $EA \wedge EB \to E(A \wedge B)$. 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 $w \in \|EA \wedge EB\|$; thus $w \in \|EA\| \cap \|EB\|$, which means that $w \in f(\|A\|, w)$ and $w \in f(\|B\|, w)$. According to condition F1 we have

(14) $\quad \|A\| \neq \emptyset \quad \text{and} \quad \|B\| \neq \emptyset,$

which implies that $f(\|A\|, w) \subseteq \|A\|$ and $f(\|B\|, w) \subseteq \|B\|$. On the other hand, it is possible that $\|A \wedge B\| = \emptyset$, which means that $w \notin f(\|A \wedge B\|, w)$. If this is the case then $\|A\| \cap \|B\| = \emptyset$; Consequently $f(\|A\|, w) \cap f(\|B\|, w) = \emptyset$.

On the other hand, if $\|A\| = \emptyset$ (or $\|B\| = \emptyset$) then $\|A \wedge B\| = \emptyset$ and so $f(\|A\|, w) = f(\|A \wedge B\|, w)$.

$A \equiv B$ iff $\|A\| = \|B\|$. In particular, if $\|A\| = \|B\| = \emptyset$, then $f(\|A\|, w) = f(\|B\|, w)$. □

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 Lindenbaum's Lemma is the following proposition.

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

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

LEMMA 10. *For every world $w \in W$ in $\mathscr{S}_c$, and every formula A, $w \vDash_{\mathscr{S}_c} A$ iff $A \in w$ (or equivalently $\|A\|^{\mathscr{S}_c} = [A]^{\mathscr{S}_c}$).*

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 $w \vDash EA$, then by the evaluation function we have $w \in f(\|A\|, w)$; by the inductive hypothesis $w \in f([A], w)$, 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 $A \not\equiv \bot$ and $[A] \neq \emptyset$; thus $f([A], w) = [EA]$ and consequently $w \in f([A], w)$. By the inductive hypothesis $w \in f(\|A\|, w)$, which implies $w \vDash EA$.

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

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 $w \vDash CA$. □

LEMMA 11. *$\mathscr{S}_c$ 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 $EA \wedge EB \to E(A \wedge B)$ is valid.

If $A \equiv \bot$ then either $f([A], w) = W - \{w\}$ or $f([A], w) = \emptyset$. In both cases $w \notin f([A], w)$. If $A \equiv \bot$, 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 $CA \neq w$, then $f([A], w) = \emptyset$, thus, trivially $f([A], w) \subseteq [A]$. □

From the two lemmata above we obtain that $\mathscr{L}_1$ is complete with respect to $\mathscr{S}$.

THEOREM 12. $\vdash_{\mathscr{L}_1} A$ *iff* $\vDash_{\mathscr{S}} A$.

7. Non-normal Worlds and Relational Models

In the previous sections we examined Elgesem's modal logic of agency and ability using semantics with different flavours. In general the selection function semantics and neighbourhood semantics give rise to the same structure: the selection function semantics focuses on the worlds where some actions can be realised 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 6 we proposed a characterisation of $\mathscr{L}_1$ based on models satisfying condition F1. According to the intended reading $f(\emptyset, w)$ is the set of worlds where the agent realises 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 worlds.[8] At any rate, the technical machinery of impossible (non-normal, queer) worlds[9] offers us the opportunity to present an alternative class of Elgesem's models for $\mathscr{L}_1$. 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, $w_\perp \Vdash A$, 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 realise the impossible, it seems plausible to have a semantic counterpart for this notion. Hansson and Gärdenfors [14] 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 the accessibility relation R in relational models) by using non-normal/impossible worlds obeying 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 [14].

Scott–Montague models were 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 $\neg C\bot$ 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 – when condition F1, which does not rule out the presence of impossible worlds in a model, is assumed – or explicitly – when the impossible world $w_\bot$ is introduced –, 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 $\mathscr{L}_1$ (and $\mathscr{L}_2$). In the rest of this section we will investigate this issue.

Classical modal logics are characterised by models with the following structure [10]:

(15) $\langle W, N, R^*, v\rangle$

where W, v 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

$$(16)\qquad w \vDash \Box A \quad \text{iff} \quad w \in N \quad \text{and} \quad \exists R \in R^* \quad \text{such that } \forall x(wRx \text{ iff } x \vDash A)$$

The set of non-normal worlds is denoted by Q (where $Q = W - N$). Alternatively we could define a model as $\langle W, Q, R^*, v\rangle$. Clearly if $w \in Q$, for any formula A, $w \nvDash \Box A$. Worlds in Q correspond 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 component. Fortunately the two operators are related by axiom A4, thus we can adopt the structure (from now on we will use X as a variable ranging over C, E)

(17) $\langle W, Q^E, Q^C, R^E, R^C, v\rangle$

where W is a set of possible worlds, Q^E and Q^C are sets 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 v 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);
R5 $\forall R \in R^C \forall w \exists x(wRx)$ (all relations in R^C are serial).

As we shall see, $\mathscr{L}_1$ is determined by the class of relational models satisfying R1–R4, and $\mathscr{L}_2$ 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 $\mathscr{N} = \langle W, N^E, N^C, v\rangle$ be a neighbourhood model and $\mathscr{R} = \langle W, Q^E, Q^C, R^E, R^C, v\rangle$ be a relational model such that*

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

Then for all formulas A: $\vDash_{\mathscr{N}} A$ iff $\vDash_{\mathscr{R}} A$.

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 we consider only the cases of the modal operators.

$$\begin{aligned} w \vDash_{\mathscr{N}} XA &\Rightarrow \|A\|^{\mathscr{N}} \in N^X_w \\ &\Rightarrow N^x_w \neq \emptyset \\ &\Rightarrow w \notin Q^X \\ &\Rightarrow \exists R \in R^X : R_w = \|A\|^{\mathscr{N}} = \|A\|^{\mathscr{R}} \\ &\Rightarrow \exists R \in R^X \forall x(wRx \text{ iff } x \vDash_{\mathscr{R}} A) \\ &\Rightarrow w \vDash_{\mathscr{R}} XA. \end{aligned}$$

For the other direction we have

$$\begin{aligned} w \vDash_{\mathscr{R}} A &\Rightarrow w \notin Q \text{ and } \exists R \in R^X \forall x(wRx \text{ iff } x \vDash_{\mathscr{R}} A) \\ &\Rightarrow R_w = \|A\|^{\mathscr{R}} = \|A\|^{\mathscr{N}} \\ &\Rightarrow R_w \in N^X_w \\ &\Rightarrow \|A\|^{\mathscr{N}} \in N^X_w \\ &\Rightarrow w \vDash_{\mathscr{N}} XA. \end{aligned}$$

□

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. Besides 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* $\mathscr{M}$ *there is an equivalent* (*finite*) *neighbourhood model* $\mathscr{N}$ *such that if* $\mathscr{R}$ *satisfies* Rn *then* $\mathscr{N}$ *satisfies* Cn (*for* $1 \leq n \leq 5$).

(2) *For every finite neighbourhood model* $\mathscr{N}$ *there is an equivalent finite relational model* $\mathscr{R}$ *such that if* $\mathscr{N}$ *satisfies* Cn *then* $\mathscr{R}$ *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 $\mathscr{R}$ 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.

Case R1 $\Rightarrow$ C1. According to the construction we have that $N_w^C = \{R_w : R \in R^C\}$, but given R1 for every w and every R, $R_w \neq W$, thus $W \notin N_w^C$.

Case R2 $\Rightarrow$ C2. Condition R2 states that for each world w, the set of the projections of the relations over w is closed under intersection. N_w^E is the set of all projections of the relations in R^E over w, thus N_w^E is closed under intersection.

Case R3 $\Rightarrow$ C3. Each $R \in R^E$ is reflexive, then for every normal world w, $w \in R_w$; by construction N_w^E is the set of all R_w. Therefore for every $X \in N_w^E$, $w \in X$. If w is a non-normal world then $N_w^E = \emptyset$ and C3 is vacuously satisfied.

Case R4 $\Rightarrow$ C4. If $w \in Q^E$ then, by construction, $N_w^E = \emptyset$, and for every set N_w^C, $N_w^E \subseteq N_w^C$. In case $w \in Q^C$, then also $w \in Q^E$, and we can repeat the previous argument. If $w \notin Q^E$ we have that $N_w^C = \{R_w : R \in R^E\}$, but, by condition 4 there is a relation $R' \in R^C$ such that $R_w = R'_w$. This implies that $w \notin Q^C$, and so $N_w^E \subseteq N_w^C$.

Case R5 $\Rightarrow$ C5. According to the construction we have that $N_w^C = \{R_w : R \in R^C\}$, but given R5 for every $w \notin Q^C$ and every R, $R_w \neq \emptyset$, thus $\emptyset \notin N_w^C$.

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

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

Then

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

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.

Case C1 $\Rightarrow$ R1. Let us suppose it does not hold. This means there is a relation $S \in R^C$ such that there is a world w that is in relation with all the worlds, i.e., $S_w = W$. Since $S \in R^C$ and $S_w \neq \emptyset$, then $w \notin Q^C$, and for some i, $S_w = \beta(w, i) = \Sigma^C_{w,i}$, but this implies that $S_w \in N^C_w$, and thus $W \in N^C_w$ when we obtain a contradiction.

Case C2 $\Rightarrow$ R2. By construction, for any $w \notin Q^X$ and $R \in R^X$, R_w corresponds to some set $R \in N^X_w$, and for every set $S \in N^X_w$ there is a relation $S \in R^X$ such that $S_w = S$. Thus for any two relations $R, S \in R^E$, $R_w, S_w \in N^E_w$. Since N^E_w is closed under intersection $R_w \cap S_w \in N^E_w$; again, by construction, there is a relation T such that $T_w = R_w \cap S_w$.

Case C3 $\Rightarrow$ R3. If $N^E_w = \emptyset$ then $w \in Q^E$ and the condition does not apply to it. Otherwise we have that each relation R^E_i is based on $\Sigma^E_{w,i}$ if $i \leq |N^E_w|$ or $\Sigma^E_{w,1}$ otherwise. In both cases condition $\Sigma^E_{w,x} \in N^E_w$ and condition C3 guarantee that $w \in \Sigma^E_{w,x}$. Thus $(w, w) \in R^E_i$, which implies that every relation $R \in R^E$ is reflexive.

Case C4 $\Rightarrow$ R4. Since $N^E_w \subseteq N^C_w$ and, by hypothesis, W is finite, $|N^E_w| \leq| N^C_w|$. By construction there are e relations R^E and c relations R^C (with $e \leq c$). By construction, for $i \leq |N^E_w|$, $\alpha(w, i) = \gamma(w, i)$, i.e., $\Sigma^E_{w,i} = \Sigma^C_{w,i}$. Notice that for each relation $R_w = \Sigma^X_{w,i}$ for some $i \in \mathbb{N}$. Hence we can conclude that for every relation $R \in R^E$ there is a relation $R' \in R^C$ such that $R_w = R'_w$.

Case C5 $\Rightarrow$ R5. Let us suppose it does not hold. This means there is a relation $S \in R^C$ such that there is a world that is in relation with no worlds, i.e., $S_w = \emptyset$. Since $S \in R^C$, then $w \notin Q^C$, and for some i, $S_w = \beta(w, i) = \Sigma^C_{w,i}$, but this implies that $S_w \in N^C_w$, and thus $\emptyset \in N^C_w$ when we obtain a contradiction. □

Due to the above procedure to generate such relational models, in the case of infinite $\mathscr{N}$ or $\mathscr{N}'$ models we would get non-enumerable infinitary relational structures. To avoid these complexities, it is sufficient to consider $\mathscr{N}$ and $\mathscr{N}'$ when they are finite. This is possible by preliminarily showing that $\mathscr{L}_1$ and $\mathscr{L}_2$ have the finite model property wrt the neighbourhood models previously defined. The fmp follows immediately from the results of Lewis [23] and [34] that every classical non-iterative modal logic has the finite model property.[10]

Clearly $\mathscr{L}_1$ and $\mathscr{L}_2$ are non-iterative thus we have the following theorem.

THEOREM 15. *$\mathscr{L}_1$ and $\mathscr{L}_2$ have the fmp.*

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

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

(1) $\vdash_{\mathscr{L}_1} A$ *iff* $\models_{\mathscr{R}_1} A$*;*
(2) $\vdash_{\mathscr{L}_2} A$ *iff* $\models_{\mathscr{R}_2} A$.

Proof. Let us consider only $\mathscr{L}_1$. From Theorem 3 we know that

$$\models_{\mathscr{N}} A \rightarrow \ \vdash_{\mathscr{L}_1} A,$$

which is equivalent to saying that

$$\nvdash_{\mathscr{L}_1} A \rightarrow \ \nvDash_{\mathscr{N}} A.$$

Since $\mathscr{L}_1$ has the finite model property, there is a finite model $\mathscr{N}_{FIN}$ and a world w in it such that

$$w \models_{\mathscr{N}_{FIN}} \neg A$$

According to Proposition 14 and the generation of the corresponding relational model

$$w \models_{\mathscr{R}_1} A$$

which implies

$$\nvDash_{\mathscr{R}_1} A$$

Then,

$$\nvdash_{\mathscr{L}_1} A \rightarrow \ \nvDash_{\mathscr{R}_1} A$$

and so

$$\models_{\mathscr{R}_1} A \iff \ \vdash_{\mathscr{L}_1} A$$

The proof for $\mathscr{L}_2$ and $\mathscr{R}_2$ is analogous. □

Here we want to propose a simple interpretation of relational models: the capability of an agent to realise a particular state A depends on 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.

8. 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 5. This amounts to saying 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 $\Box$ 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 $\Box$. On the contrary the neighbourhood semantics assumes two separate but related modal operators.

It is true that selection function models are widely used in conditional logics and with different evaluation clauses for modal operators they are just a "notational" variant of neighbourhood models (indeed they are both classified as Scott–Montague semantics). However Elgesem models are not standard in modal logic; the particular evaluation clauses for modal operators make Elgesem semantics different from standard selection function semantics for modal logic. Accordingly the construction of canonical models requires some ingenuity, and it is not a straightforward extension of standard construction of canonical models – in particular when the condition F1 is involved. As we have seen this condition is used to discriminate Elgesem models rejecting $\neg C\bot$ from models validating it. But this condition is relevant not only for this axiom but it is entangled with some other conditions. The proofs of soundness and completeness (Theorems 7 and 10) make essential use of this condition in the case of agglomeration for E. The axioms are clearly independent, but the semantic conditions are, in a certain sense, entangled together by the use of the same selection function to characterise two independent but related modal operators. Interestingly the connection is broken when we explicitly introduce non-normal worlds in Elgesem model to characterise $\mathscr{L}_2$ (see the introduction to Section 7). Moreover condition F1 makes clear the implicit need of non-normal worlds.

[25], among others, argues that an agent can carry out an action successfully if she has the ability as well as the opportunity do to it. Indeed Elgesem studies the relationships between ability and agency, and he correctly realises that agency implies opportunity, i.e., $EA \rightarrow \text{Opportunity } EA$. But the notion of opportunity is given in terms of agency, i.e., Opportunity $A \equiv (E\neg A \vee A)$. 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. On the contrary 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, as we said, Elgesem's semantics leads to the introduction (either implicitly or explicitly) of non-normal worlds. This can be shown if we compare the reasons why the schema $\neg C\bot$ is valid in Elgesem's logic to the conditions we have to impose to regain completeness, in this semantics, and not to adopt such as schema. As we have seen, this last is valid in Elgesem's original semantics simply because from $f(X, w) \subseteq X$ and $\|\bot\|^{\mathscr{E}} = \emptyset$, we get $f(\|\bot\|^{\mathscr{E}}, w) = \emptyset$. Recall that, intuitively, $f(X, w)$ is interpreted as the set of worlds where the agent realises the ability she has in w to bring about X. Thus agency and ability refer to the same set of worlds selected by f: if it is impossible to have the ability to achieve $\bot$ so it is similarly impossible to achieve it. If we look at the conditions that allow us to regain completeness without accepting (7) we realise that condition F1 confines the inclusion $f(X, w) \subseteq X$ only to the cases where $X \neq \emptyset$; otherwise, $w \notin f(X, w)$. Notice that, strictly speaking, this condition does not exclude that there is one world w where $\bot$ holds, but simply that, if such a world exists, it is not picked out by $f(X, w)$. This w, if it exists, is nothing but an impossible world; therefore if there is a second world x such that $w \in f(\emptyset, x)$, then x is a non-normal world. This idea seems to be confirmed by focusing on another alternative. In fact, the validity of (7) in Elgesem's original semantics results from the condition $f(X, w) \subseteq X$, applied to any X, and from using a single selection function for both C and E. Thus, let us suppose to define two different functions f_E and f_C as in standard Scott–Montague selection function models. Following Elgesem, we will simply state that $f_E(X, w) \subseteq X$ for any X, and, in virtue of (6), that $f_E(X, w) \subseteq f_C(X, w)$. As in Elgesem's approach, we will obtain $f_E(\|\bot\|^{\mathscr{E}}, w) = \emptyset$, but this will not imply that the same holds for f_C. This means that schema (7) is not required and that the set of worlds picked out by f_C for a world w may include a world x where an agent is able to do $\bot$; hence w is a non-normal world, as we argued. To sum up, the semantical implication we have to accept if we want to adopt Elgesem's strategy – using one selection function to represent E and C and having full reflexivity for E – is that of excluding non-normal worlds: we have necessarily to adopt (7). Otherwise, we may change strategy but this requires (either implicitly or explicitly) the existence of non-normal worlds. The first option is of course sound, as we argumented in Section 4 with regard to the idea of practical agency. However, it is not the only available because different interpretations may be assigned to the idea of agency.

On the other hand, as we argued, the introduction of non-normal worlds within Elgesem's semantics is not the best choice as their interpretation is not fully 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. This is the advantage of multi-relational semantics, which allows to avoid that non-normal worlds are like black-boxes, as we said, namely entities without any further analysis of their (internal) structure. This semantics, on the other hand, does not lend itself to easy manipulation and calculation. Neighbourhood models are quite the opposite: they are easier to manipulate and work with for non-iterative logic, even though they are not particularly intuitive. Both aspects are mainly due to their close relationships with algebraic semantics for modal logic. Elgesem's semantics, at the end, is simple and provides an intuitive interpretation, but it is not transparent to extensions. Hidden semantic relationships can appear when such a semantics is extended: two examples we have seen here are the role played by reflexivity in making $\mathscr{L}_1$ incomplete, and the relationships between $\neg C\bot$ and agglomeration for E.

ACKNOWLEDGEMENTS

Preliminary versions of this paper were presented at the Annual Meeting of the Australasian Association for Logic (Dunedin, New Zealand, 17–18 January 2004) and at Advances in Modal Logic 2004 (Manchester, UK, 9–11 September 2004). We would like to thank all anonymous referees for their valuable comments and suggestions that improved the presentation of this paper.

The first author was supported by Australia Research Council under Discovery Project No. DP0452628 on "Combining modal logic for dynamic and multi-agents systems".

NOTES

[1] As 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., [19]). Thus the first is represented by the operator E, while the second by C. Of course, both are labelled by agents as well.

[2] According 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 (7). 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 [9]. This second aspect will not be considered here, since it does not seem relevant with regard to the aims of this paper.

[3] Elgesem's semantics for the modal logic of agency and ability is a structure $\langle W, f_1, \ldots, f_n, V\rangle$ (cf. [9, p. 20] and [8, p. 54]), where each f_i, $1 \leq i \leq n$, is a function as in (8) 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 accidence. 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.

[4] In fact, the general idea of avoidability is often linked with rejecting the ability to realize the impossible. This view is maintained, in a way, by Mark Brown in [4, p. 18], where the author argues that the ability to realise $\perp$ makes sense only in bizarre-sounding examples in which, for any goal A, the agent, though able to do A, is prevented to realise it for all possible situations.

[5] The notions at hand in this interpretation are closely related to the long-standing problem of the validity of norms in a normative system; for a logical account, see, among others, the seminal work by Alchourrón and Bulygin [1].

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

[7] Elgesem [8] claims that his logic is complete. However the proof is only sketched. To the best of our knowledge this paper provides the first full proof of completeness for $\mathscr{L}_1$ and $\mathscr{L}_2$ with respect to Elgesem models.

[8] It 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. In particular we assume that impossible worlds are worlds where $\perp$ is true, however they are impossible according to a classical reading: thus, if $\perp$ is true everything else is true, and a fortiori $\top$ is true. Thus it is not possible to use $\top$ to discern possible worlds from impossible ones.

[9] Non-normal (or queer) worlds were introduced by Kripke in [22] to give a possible world semantics to Lewis' systems S2 and S3. Since then they have been used to provide models for several non-normal and intensional logics.

[10] A modal logic is non-iterative iff it can be axiomatised 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.

REFERENCES

1. Alchourrón, C. E. and Bulygin, E.: *Normative Systems*, Springer-Verlag, Wien, 1971.
2. Belnap, N. and Perloff, M.: Seeing to it that: A canonical form for agentives, *Theoria* **54** (1988), 175–199.
3. Belnap, N. and Perloff, M.: The way of the agent, *Studia Logica* **51** (1992), 463–484.
4. Brown, M. A.: On the logic of ability, *J. Philos. Logic* **17** (1988), 1–26.
5. Carmo, J. and Pacheco, O.: Deontic and action logics for organized collective agency modeled through institutionalized agents and roles, *Fund. Inform.* **48** (2001), 129–163.
6. Chellas, B.: *The Logical Form of Imperatives*, Perry Lane Press, Palo Alto, 1969.
7. Chellas, B.: *Modal Logic: An Introduction*, Cambridge University Press, Cambridge, 1980.
8. Elgesem, D.: Action theory and modal logic, PhD, Institut for filosofi, Det historisk-filosofiske fakultetet, Universitetet i Oslo, 1993.
9. Elgesem, D.: The modal logic of agency, *Nordic J. Philos. Logic* **2**(2) (1997), 1–46.
10. Gasquet, O. and Herzig, A.: From classical to normal modal logic, in H. Wansing (ed.), *Proof Theory of Modal Logic*, Kluwer, Dordrecht, 1996, pp. 293–311.
11. Gelati, J., Governatori, G., Rotolo, A. and Sartor, G.: Declarative power, representation, and mandate: A formal analysis, in T. Bench-Capon, A. Daskalopulu and R. Winkels (eds.), *Legal Knowledge and Information Systems*, IOS Press, Amsterdam, 2002, pp. 41–52.
12. Governatori, G., Gelati, J., Rotolo, A. and Sartor, G.: Actions, institutions, powers. Preliminary notes, in G. Lindemann, D. Moldt, M. Paolucci and B. Yu (eds.), *International Workshop on Regulated Agent-Based Social Systems: Theories and Applications (RASTA'02)*, Fachbereich Informatik, Universität Hamburg, 2002, pp. 131–147.
13. Governatori, G. and Rotolo, A.: Defeasible logic: Agency, intention and obligation, in A. Lomuscio and D. Nute (eds.), *Deontic Logic in Computer Science*, LNAI 3065, Springer-Verlag, Berlin, 2004, pp. 114–128.
14. Hansson, B. and Gärdenfors, P.: A guide to intensional semantics, in *Modality, Morality and Other Problems of Sense and Nonsense. Essays Dedicated to Sören Halldén*, Gleerup, Lund, 1973, pp. 151–167.
15. Hilpinen, R.: On action and agency, in E. Ejerhed and S. Lindström (eds.), *Logic, Action and Cognition: Essays in Philosophical Logic*, Kluwer Academic Publishers, Dordrecht, 1997, pp. 3–27.
16. Hintikka, J.: Impossible possible worlds vindicated, *J. Philos. Logic* **4** (1975), 475–484.
17. Horty, J. F. and Belnap, N.: The deliberative stit. A study of action, omission, ability, and obligation, *J. Philos. Logic* **24** (1995), 583–644.
18. Jones, A. J. I. and Sergot, M.: A formal characterisation of institutionalised power, *J. IGPL* **3** (1996), 427–443.
19. Jones, A. J. I.: A logical framework, in J. Pitt (ed.), *Open Agent Societies: Normative Specifications in Multi-Agent Systems*, Wiley, Chichester, 2003, Chapter 3.
20. Kenny, A.: *Will, Freedom and Power*, Blackwell, Oxford, 1975.
21. Kenny, A.: Human abilities and dynamic modalities, in J. Manninen and R. Tuomela (eds.), *Essays on Explanation and Understanding*, Reidel, Dordrecht, 1976.
22. Kripke, S. A.: Semantical analysis of modal logic II, in Addison, Henkin and Tarski (eds.), *The Theory of Models*, North-Holland, Amsterdam, 1965, pp. 206–220.

23. Lewis, D.: Intensional logic without iterative axioms, *J. Philos. Logic* **3**(4) (1974), 457–466.
24. Meyer, J.-J. Ch. and Wieringa, R. J.: Deontic logic: A concise overview, in J.-J. Ch. Meyer and R. J. Wieringa (eds.), *Deontic Logic in Computer Science: Normative System Specification*, Wiley, New York, 1993, pp. 3–16.
25. Padmanabhan, V. N., Governatori, G. and Sattar, A.: Actions made explicit in BDI, in M. Stumptner, D. Corbett and M. J. Brooks (eds.), *Advances in Artificial Intelligence*, LNCS 2256, Springer-Verlag, 2001, pp. 390–401.
26. Pörn, I.: *The Logic of Power*, Blackwell, Oxford, 1970.
27. Pörn, I.: *Action Theory and Social Science: Some Formal Models*, Reidel, Dordrecht, 1977.
28. Santos, F. and Carmo, J.: Indirect action. Influence and responsibility, in M. Brown and J. Carmo (eds.), *Deontic Logic, Agency and Normative Systems*, Springer, Berlin, 1996.
29. Santos, F., Jones, A. J. I. and Carmo, J.: Action concepts for describing organised interaction, in *Thirtieth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Los Alamitos, 1997.
30. Segerberg, K.: *An Essay in Classical Modal Logic*, Filosofiska Studier 13, Uppsala Universitet, Uppsala, 1971.
31. Segerberg, K.: Bringing it about, *J. Philos. Logic* **18** (1989), 327–347.
32. Segerberg, K.: Getting started: Beginnings in the logic of action, *Studia Logica* **51** (1992), 347–358.
33. Sergot, M. and Richards, F.: On the representation of action and agency in the theory of normative positions, *Fund. Inform.* **48** (2001), 273–293.
34. Surendonk, T. J.: Canonicity for intensional logics without iterative axioms, *J. Philos. Logic* **26**(4) (1997), 391–409.

GUIDO GOVERNATORI
School of Information Technology and Electrical Engineering,
The University of Queensland,
Brisbane, QLD 4072, Australia
e-mail: guido@itee.uq.edu.au

ANTONINO ROTOLO
CIRSFID, Law Faculty, University of Bologna,
Via Galliera 3, 40121, Bologna, Italy
e-mail: rotolo@cirfid.unibo.it