A Modal Logic of Intentional Communication

Marco Colombetti

Artificial Intelligence and Robotics Project
Department of Electronics and Information Sciences
Politecnico di Milano

Abstract. I propose a propositional modal logic of intentional communication, a circular concept which develops an idea by Airenti, Bara and Colombetti (1993). A communication operator is added to a multi-modal language of individual belief, common belief, and intention, and its possible world semantics is justified through a fixpoint construction. A normal modal system for the communication operator is defined, and shown to be sound and complete. Within this framework, I prove some properties of intentional communication and give sufficient conditions for an action to be communicative.

To appear on *Mathematical Social Sciences.*

Send correspondence to the author, Department of Electronics and Information Sciences, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy. Voice: (39) (2) 2399 3525. Fax: (39) (2) 2399 3411. Email: `Marco.Colombetti@polimi.it`

1. Introduction

In the last few years, the concept of common belief[1] has received a great deal of attention in such diverse fields as economics, decision theory, social psychology, cognitive science, artificial intelligence and distributed computer science. Recently, several technical papers have proposed sound and complete axiomatizations of this concept as modal logical systems, with either Kripkean or neighborhood semantics (Halpern and Moses, 1992; Lismont and Mongin, 1995; Bonanno, 1996).

To assume that a group of agents have common belief of some proposition is a strong condition, because it requires that each agent believes the proposition, that each agent believes that each agent believes the proposition, and so on *ad infinitum*. It is therefore important to analyze the situations which can bring common beliefs into existence. Typically, common beliefs are shown to arise when some proposition is *public*, that is, when it cannot hold without all agents believing that it holds. This treatment appears to suit common beliefs that arise because some fact is intrinsically so evident that it would be impossible for a group of rational agents not to recognize it. For example, if a group of people have dinner together around a table, we can safely assume that the fact they are dining together is common belief of the group.

In reality, however, many common beliefs do not arise from this kind of situations, but are the product of a process of communication within a group of agents. Suppose, for example, that Alice (the *speaker*) tells all other members of a group (the *audience*) that she feels tired. Then, under appropriate conditions, two facts (among others) become common beliefs of the group: (i) that Alice communicated that she is tired, and (ii) that Alice is tired. The most natural justification for (i) is that to communicate is a public event; however, the same kind of justification does not apply to (ii), because being tired is typically not public. In fact, (ii) has become public as an effect of Alice's communicative act. It seems therefore that the notion of communication deserves accurate analysis; ideally, such an analysis should result into a proper extension of some suitable epistemic logic. In this paper, I propose a treatment of communication in an extension of a frequently adopted modal logic of belief, that is, the normal multi-modal system $KD45_n$.

Recently, a number of scientists (notably Cohen and Levesque, 1990) have tried to develop a treatment of communication within a general formal theory of action. However, building such a theory is a formidable task, which so far has not been accomplished in a satisfactory way. The approach advocated in this paper is different: keeping an eye on Occam's razor, I shall try to do as much as I can with simple technical tools. My hope is that it will be possible to embed the results obtained in this way into wider logical treatments of human action.

Communication has a large variety of different facets, and it is unlikely that a single definition can be provided that covers all the reasonable acceptations of the term. It is not at all obvious, for example, that communication among humans, among nonhuman animals, and

---

[1] The relevant literature deals with both common belief and common knowledge. The two concepts are strictly related, in that knowledge is typically identified with true belief. In this paper I shall only be concerned with belief.

among artificial systems can be defined in a similar fashion. In this paper I shall only be concerned with communication among human beings, which is evidently the most relevant case for the social sciences.

As remarked by Sperber and Wilson (1986, Chapter 1), so far human communication has been treated from two different perspectives: the *code model* of information theory (Shannon and Weaver, 1949), and the *inferential model,*[2] based on ideas introduced by analytical philosophers like Grice (1957), Strawson (1964), Lewis (1969), and Schiffer (1972). The code model is basically concerned with the transmission of information from a source to a receiver. It can be used, for example, to study possible loss of information through a channel, but it is not the right tool to analyze the effects of communication on an agent's belief system. Models of the inferential type, on the contrary, regard communication as a kind of intentional activity carried out by an agent to achieve an effect on an audience, and appear to approach the matter from the right side, in view of my present goals. Therefore, in this paper I shall only be concerned with the inferential model.

Since Grice's influential paper on meaning (1957), it is widely accepted that communicative acts involve higher-order intentions. That is, communication not only encompasses a speaker's first-order intention, I^1, to achieve an effect on an audience, but also the speaker's second-order intention, I^2, that I^1 be recognized by the audience. The second-order intention is meant to capture the overt nature of communication, allowing one to distinguish a genuine communicative act from an attempt to achieve an effect on a group of people in a concealed way. Suppose for example that Alice wants Bob to believe that she is at home. Alice can obtain this effect through various means, and in particular:

(i) by turning on the light and the TV set in her sitting room; or

(ii) by calling Bob on the phone and telling him "I am at home".

In case (i), Alice has the first-order intention, I^1, to make Bob believe that she is at home, but need not have the second-order intention, I^2, to let Bob know that she has intention I^1; however, if Alice does not hold intention I^2, her act cannot be properly described as communication. In case (ii), on the contrary, by telling Bob that she is at home, Alice is intentionally making her first-order intention known to him; according to Grice's viewpoint, this can be regarded as a proper instance of communication.

Strawson (1964) was probably the first to remark that second-order intentions are not sufficient to define communication, and proposed to consider the third-order intention, I^3, that also I^2 be recognized. Examples showing why it should be so are fairly baroque, but I shall try to elaborate case (i) above to make Strawson's standpoint clear. Suppose that Alice actually intends Bob to recognize her intention to make him believe that she is at home (this is the second-order intention, I^2), without overtly telling him so. In an appropriate context, Alice might make sure that Bob sees her turn on the light and the TV set, behaving in such a way that Bob will suspect, or even firmly believe, that Alice intends him to believe that she is at home. In such a case, Alice would actually entertain intention I^2 (that I^1 be recognized by Bob), but still she would not communicate to Bob that she is at home. Indeed, this

[2] This term is, in my opinion, rather misleading, because it suggests that the process of communication is itself an inferential process. While Sperber and Wilson endorse this view, I think that the fact that a model of a process is inferential (i.e., expressed as a logical system) does not imply that the process itself be inferential in any meaningful way.

counterexample to Grice's definition is ruled out by Strawson's requirement that Alice has the third-order intention, I^3, that I^2 be recognized by Bob.

The function of Strawson's third-order intention is to guarantee that a communicative act is truly overt. However, once we start climbing up the hierarchy of intentions, there seems to be no reason to stop at level three. In fact, Airenti, Bara and Colombetti (1993) argue that no finite conjunction of finite-order intentions can do the job, because (page 206):

> "… if an *n*th-order intention is required in the definition [of communication], then the actor might fail to entertain the intention of order n+1. The interactive situation is therefore not fully overt, because a part of it is not meant to be recognized, but rather is kept private by the actor."

From a technical point of view, this implies that either one postulates an infinite hierarchy of intentions up to order ω, or provides a circular definition of communication: a situation that is already familiar to logicians from the analysis of common belief.

An iterative analysis of communication is proposed by Perrault (1990) within the framework of default logic. However, the iterative approach (like in the case of common belief) does not allow one to capture the notion of communication in finitary terms, and forces one to deal with it at the metalinguistic level. The result is that reasoning about communication cannot take place within the logic.

A circular approach to communication is adopted by Airenti, Bara and Colombetti (1993), who do not directly define a notion of communication, but rather introduce a circular concept of *communicative intention* through an appropriate "fixpoint axiom". Their treatment, however, has a number of technical shortcomings. The first is the system proposed by Airenti *et al.* is not fully formal, and in particular is not endowed with formal semantics. The second is that the fixpoint axiom, as we shall see, does not completely characterize communicative intention. The third is that the definition of intentional communication appears to be more general and flexible if one separates the intentional component from the communicative one.

In the rest of this article, I propose a formal treatment of communication within a sound a complete normal modal logic. A difficulty for theories that deal with circular states, like common belief and (in this paper) communication, is to provide conditions under which such states hold. Circular states require very strong conditions to occur; however, in all concrete cases the communicating agents will ultimately have to rely on their individual beliefs to judge whether something has been communicated. At the end of Section 6, I propose a solution to this problem.

The plan of the article is the following. In Section 2, I introduce a circular definition of communication at the intuitive level. In Section 3, I define a basic logic of individual beliefs and intentions, which is extended to cover common belief in Section 4. In Section 5, I give a formal definition of communication. Some properties of communication are then proved in Section 6, and a short discussion is provided in Section 7. The Appendix reports the completeness proof for the modal logic of communication proposed in Section 5.

2. The Basic Intuition

Literally, to communicate means to make something common. I shall take this etymology seriously, and regard communication as an intentional activity resulting into a common belief. More precisely, let $A = \{1, ..., n\}$ be a finite set of n agents, and let us suppose that an agent, a, communicates something to the other n–1 agents of A; I shall designate a as the *speaker*, and the other n–1 agents as the *audience*.[3] For a given proposition φ, I take that a successfully communicates that φ if and only if:

- φ is a common belief of A;
- that a intends to communicate that φ is also a common belief of A.

Let us use the following notation (rather informally, for the time being):

- $B_*\varphi$, to mean that φ is a common belief of A;
- $I_a\varphi$, to mean that agent a intends to perform φ;
- $C_a\varphi$, to mean that agent a communicates that φ to the audience $A-\{a\}$.

Then, intuitively, communication should be defined so that the following scheme is valid:

$$(\mathrm{F}_C) \qquad C_a\varphi \equiv B_*(\varphi \wedge I_aC_a\varphi).$$

F_C is formally related to the well-known fixpoint axiom of mutual belief, and I shall therefore refer to it as the *fixpoint axiom of communication*. I shall refer to the proposition expressed by φ in $C_a\varphi$ as to the *content* of communication.

The intuitive definition of communication via F_C is bound to be controversial. For example, it can be argued that F_C is a very strong condition, because it requires the speaker to be *sincere* about φ (i.e., the speaker must believe that φ), and the audience to be *convinced* by the speaker's communicative act. Indeed, one might argue for a weaker definition, for example by assuming that successful communication coincides with the common belief that the speaker intends to communicate that φ (in the previous sense), without requiring that common belief that φ be also achieved. Let us write $C'_a\varphi$ for this weaker notion of communication; then we have the fixpoint axiom:

$$C'_a\varphi \equiv B_*(I_aB_*\varphi \wedge I_aC'_a\varphi).$$

That C'_a is weaker than C_a is reflected by the fact that (assuming rather obvious logical properties for our operators) $C_a\varphi$ entails $C'_a\varphi$, but not *vice-versa*. Establishing which of the two notions of communication is more basic is certainly a relevant philosophical question. However, from the technical point of view taken in this article, the question seems of secondary importance. In fact, it turns out that the two concepts are mutually interdefinable through the following schemes:

$$C'_a\varphi \equiv B_*I_aC_a\varphi,$$
$$C_a\varphi \equiv B_*\varphi \wedge C'_a\varphi.$$

Another plausible definition of communication can be based on the idea that successful communication that φ, besides implying the common belief of the intention to communicate that φ, does not necessarily imply the common belief that φ, but only the common belief that *the speaker believes* that φ:

---

[3] It is important to remark that by using this terminology I am not assuming that communication only involves (or even partially involves) the use of speech. My treatment applies to communication based on any type of code, either verbal or nonverbal.

$$C''_a\varphi \equiv B_*(B_a\varphi \wedge I_aC''_a\varphi).$$

It is easy to check that, again assuming obvious logical properties for the operators, C''_a is related to C_a through:

$$C''_a\varphi \equiv C_aB_a\varphi.$$

Within the logic of communication proposed in Section 5, C''_a is strictly weaker than C_a, because $C_a\varphi \supset C_aB_a\varphi$ is a theorem, while the converse implication is not.

In my opinion, the strongest definition of communication, $C_a\varphi$, is preferable, because it is formally simpler and it expresses in a more direct way the idea that the basic reason for communicating is to make some content common in a group of agents. However, a treatment similar to the one proposed in this article can be developed also for different concepts of communication, if they are defined through fixpoint axioms analogous to F_C.

All definitions previously discussed are based only on common belief and individual intention. Indeed, different approaches to communication are possible. One might argue that the reason for communicating something is not that the content of communication becomes common belief, but rather that the speaker becomes *committed* to such a content in front of the audience. For example, one may regard the utterance "I am tired" not as the speaker's attempt to convince the audience that she is tired, but rather as an action that commits the speaker to carry on her interaction with the audience *as if* she were tired. However, I shall not pursue this approach here, because it presupposes a suitable formalization of the notion of commitment, which lies outside of the scope of this article.

Another possible objection to my view of communication is that it might seem adequate only for assertive communicative acts, that is, acts that communicate that something is the case. The utterance "I am tired" conforms to this type, but what about questions, requests, promises, an so on? Consider for example the request "Close the door", made by a to a second individual b. If we take the content of the request to be

$$\varphi = \text{'}b \text{ closes the door'},$$

then F_C does not discriminate the request from a communicative act in which a simply asserts that b is closing the door. This problem, however, has to do not with communication *per se*, but with what we understand its content to be. It appears that to accommodate nonassertive communicative acts it is necessary to include in what I have called content also the *illocutionary force* of a speech act (Austin, 1962; Searle, 1969), that is, the fact that it counts as an assertion, or a request, etc. Again, a detailed development of this point is beyond the scope of the present article (but see Airenti, Bara and Colombetti, 1993).

In the following sections I show that starting from the intuition underlying F_C it is possible to build a modal logic of communication. More precisely, I show that given a suitable modal logic of individual beliefs and intentions, one can extend it to a sound and complete modal logic of communication. The extension is obtained by adding new modal operators to the formal language, and by relating them to the original operators through axioms and inference rules. The circularity of the notion of communication will be reflected in the structure of such axioms and inference rules, which in turn correspond to a suitable construction on the semantic side. Thus, the definition of communication will be technically similar to the by now well-known circular definition of common belief.

My workplan presupposes the development of a modal logic of individual beliefs and intentions. While various options are available for belief, no generally accepted logic of intention exists yet, and this paper is not going to change the situation. Luckily, the analysis of communication can be founded only on very general properties of intention; it is thus unnecessary to rely on a full-fledged theory of such a mental state in order to gain sufficient insight into communication. In particular, I shall not deal with aspects involving time, and shall provide a barely minimal treatment of the inter-relationships between beliefs and intentions.

3. Belief and Intention

The formal language L_0

To define the basic logic of belief and intention, BI_0, let us start from a propositional modal language L_0 based on a finite or denumerable set, $P = \{p, q, r, ...\}$, of propositional constants, on the propositional connectives ~ and $\wedge$, and on two families of modal operators, $(B_a)_{a\in A}$ and $(I_a)_{a\in A}$, where $A = \{1, ..., n\}$ is a finite set of agents. Sentences in L_0 will be denoted by the small Greek letters φ, ψ, χ; parentheses will be used as usual to clarify the structure of sentences.

The most intuitive interpretations of $B_a\varphi$ and $I_a\varphi$ are that agent a believes that φ and intends that φ, respectively. Given this reading, it might look odd that the operator I_a is applied to arbitrary sentences; it seems that not any proposition, but only propositions that describe actions performed by a, can legitimately be the object of a's intentions. It is important, however, not to be led astray by the intuitive reading: the only correct interpretation of modal sentences is the one enforced by the formal semantics of L_0. As we shall see below, $B_a\varphi$ holds if and only if φ is true in all states of affairs compatible with a's current beliefs, and $I_a\varphi$ holds if and only if φ is true in all states of affairs compatible with a's current intentions. With such an interpretation, it becomes perfectly acceptable to apply an intention operator to arbitrary sentences. For example, if φ means that the door is closed, $I_a\varphi$ should be read as 'a intends to act in such a way that he or she brings about a state of affairs in which the door is closed'. More simply and idiomatically, I shall usually rephrase this as 'a intends to close the door'; however, we should always keep the correct reading in mind.

Semantics of L_0

For language L_0, let us define a Kripkean semantics. A *frame* for L_0 is an $(2n+1)$-tuple $F = \langle W,(\mathcal{B}_a)_{a\in A},(\mathcal{I}_a)_{a\in A}\rangle$, where $W = \{w, w', w'', ...\}$ is a nonempty set of possible worlds, and $\mathcal{B}_a$, $\mathcal{I}_a \subseteq W^2$, for $a \in A$, are accessibility relations. Intuitively, possible worlds are comprehensive states of affairs, $w\mathcal{B}_a w'$ means that w' is compatible with a's beliefs at w, and $w\mathcal{I}_a w'$ means that w' is compatible with a's intentions at w.

A *model* for L_0 is a pair $M = \langle F,v\rangle = \langle W,(\mathcal{B}_a)_{a\in A},(\mathcal{I}_a)_{a\in A},v\rangle$, where F is a frame and the valuation v: $P \rightarrow \mathcal{P}(W)$ is a function assigning a proposition[4] $v(p) \subseteq W$ to each propositional constant $p \in P$. Given a model M, we can recursively define an interpretation function $[\![-]\!]_M$, which assigns a proposition to each sentence:

---

4 A *proposition* is defined extensionally as an arbitrary subset of W.

$$
\begin{aligned}
[\![p]\!]_M &= v(p), \text{ for all } p \in P,\\
[\![\sim\varphi]\!]_M &= W - [\![\varphi]\!]_M,\\
[\![\varphi \wedge \psi]\!]_M &= [\![\varphi]\!]_M \cap [\![\psi]\!]_M,\\
[\![B_a\varphi]\!]_M &= \{w \in W: wB_aw' \Rightarrow w' \in [\![\varphi]\!]_M\},\\
[\![I_a\varphi]\!]_M &= \{w \in W: w\mathcal{I}_aw' \Rightarrow w' \in [\![\varphi]\!]_M\}.
\end{aligned}
$$

The connectives $\vee$, $\supset$, and $\equiv$ are introduced as abbreviations in the customary way.

As usual, I say that φ *is true in a model M at world w*, in symbols $M,w \models \varphi$, if and only if $w \in [\![\varphi]\!]_M$; that φ *is true in a model M*, in symbols $M \models \varphi$, if and only if $M,w \models \varphi$ for all $w \in W$; and that φ *is valid* (with respect to a reference class of models $\mathcal{M}$), in symbols $\models \varphi$, if and only if $M \models \varphi$ for all models $M \in \mathcal{M}$.

The minimal normal system for L_0

The class of all models (with no restrictions on the underlying frames) determines[5] the minimal normal system based on following axioms and inference rules::

(P) *All propositional tautologies*,

(K_B) $B_a\varphi \wedge B_a(\varphi \supset \psi) \supset B_a\psi$,

(K_I) $I_a\varphi \wedge I_a(\varphi \supset \psi) \supset I_a\psi$,

(RMP) $\dfrac{\varphi, \quad \varphi \supset \psi}{\psi}$,

(RN_B) $\dfrac{\varphi}{B_a\varphi}$,

(RN_I) $\dfrac{\varphi}{I_a\varphi}$.

The adoption of a normal logic for both belief and intention needs some justification. As is well known (see for example Chellas, 1980), if a modal operator O has a normal logic, it has the *monotonicity property*, expressed by the following derived inference rule:[6]

(RM_O) $\dfrac{\varphi \supset \psi}{O\varphi \supset O\psi}$.

When O is the belief operator B_a, monotonicity brings in the famous problem of *logical omniscience*. Many proposals have been put forward with the aim of avoiding this difficulty, but none has gained universal acceptance. In this article, I shall not further discuss this matter.

---

5 A class of models $\mathcal{M}$ is said to *determine* a system Σ if and only if Σ is sound and complete with respect to $\mathcal{M}$.

6 For a formal system Σ, a *derived inference rule* is an inference rule which is not part of the definition of Σ, but can be added to Σ without properly extending the set of theorems.

The adoption of a normal logic for intention is even more controversial. In fact, monotonicity of intention is judged to be unacceptable by several authors (see for example Konolige and Pollack, 1993). The typical argument goes like this. Suppose that a dentist intends to drill a patient's molar. Given that drilling a molar implies causing pain, we can derive that the dentist intends to cause pain, which in general is not true. Therefore, normal systems are not adequate to formalize the concept of intention.

The argument, however, is flawed. The fact that drilling molars causes pain is not a logical theorem, but just a truth relative to some particular models, as it is logically possible to drill a molar without causing any pain. To stress this point, consider a genuine logical theorem: since a molar is a type of tooth, drilling a molar implies drilling a tooth. Any normal logic of intention will allow us to derive that if a dentist intends to drill a molar, then he or she intends to drill a tooth: but this is correct.

Another possible objection regards the rule of necessitation. Even if it is acceptable to derive that any agent believes in all theorems, it seems awkward to assume that any agent *intends* all theorems. As I have already remarked, however, we should be very careful with our reading of $I_a\varphi$. What the rule of necessitation really says is that all theorems hold at all states of affairs that an agent might intend to bring about: again, an entirely acceptable assumption.

The system BI_0

Let me now proceed to the more specific axioms for belief and intention. Following a well-established tradition, I assume the logic of each B_a operator to be $\mathrm{KD45}_n$. This means that the following axioms hold:

(D_B) $\quad B_a\varphi \supset {\sim}B_a{\sim}\varphi$,

(4_B) $\quad B_a\varphi \supset B_aB_a\varphi$,

(5_B) $\quad {\sim}B_a\varphi \supset B_a{\sim}B_a\varphi$.

These axioms respectively express the assumptions of coherence, positive introspection and negative introspection of belief. I assume the same properties for intention:[7]

(D_I) $\quad I_a\varphi \supset {\sim}I_a{\sim}\varphi$,

(4_{IB}) $\quad I_a\varphi \supset B_aI_a\varphi$,

(5_{IB}) $\quad {\sim}I_a\varphi \supset B_a{\sim}I_a\varphi$.

Axiom D_I expresses a reasonable coherence requirement on intentions: an agent cannot intend both φ and ${\sim}\varphi$ at the same time; this is one of the properties which distinguish intentions from weaker volitional mental states, like desires. As regards Axioms 4_{IB} and 5_{IB}, they imply a complete introspective belief of the agent's intention; if we identify belief with awareness, the possibility of unconscious intentions is ruled out. I would, however, suggest that we resist this temptation: as defined through modal logic, belief is a very abstract concept, and its identification with the psychological state of awareness is not licensed.

To summarize, the normal system BI_0 is defined by the axioms:

---

7 Note that by positive (negative) introspection of intention I mean that positive (negative) intentions can be imported into the context of a *belief* operator.

P, K_B, D_B, 4_B, 5_B, K_I, D_I, 4_{IB}, and 5_{IB},

and by the inference rules:

RMP, RN_B, and RN_I.

Theorems of BI_0 will be denoted by $\vdash \varphi$.

Correspondences and determination

All specific axioms of BI_0 have first-order correspondences (van der Hoek, 1993), which I list below (free variables in the correspondence formulae are universally quantified over *W*):

axiom	*correspondence*	
D_B	B_a is *serial*:	$\exists w' . wB_aw'$;
4_B	B_a is *transitive*:	wB_aw' & $w'B_aw'' \Rightarrow wB_aw''$;
5_B	B_a is *Euclidean*:	wB_aw' & $wB_aw'' \Rightarrow w'B_aw''$;
D_I	I_a is *serial*:	$\exists w' . wI_aw'$;
4_{IB}	I_a is *transitive over* B_a:	wB_aw' & $w'I_aw'' \Rightarrow wI_aw''$;
5_{IB}	I_a is *Euclidean over* B_a:	wB_aw' & $wI_aw'' \Rightarrow w'I_aw''$.

As usual, I rely on correspondences to define the reference class of models: the class of models for BI_0, denoted by $\mathcal{BI}$, is made up by all models whose underlying frames satisfy all the above conditions. Validity with respect to $\mathcal{BI}$ will be denoted by $\models \varphi$.

It is easy to see that the class of models $\mathcal{BI}$ determines the system BI_0. Given the above correspondences BI_0 is obviously sound. Moreover, van der Hoek (1993) has proved that the six specific axioms I have adopted for BI_0 are *canonical*; that is, they are true in all models built by adding arbitrary valuations to the canonical frame.[8] This is sufficient to guarantee that the canonical model of BI_0 belongs to $\mathcal{BI}$, and therefore that BI_0 is complete.

A few basic properties of belief and intention

Only a few theorems of BI_0 are relevant to the rest of this paper. The first two such theorems tell us that an agent entertains a belief (intention) if the agent believes that he or she entertains it; in symbols

$\vdash B_aB_a\varphi \supset B_a\varphi$,
$\vdash B_aI_a\varphi \supset I_a\varphi$.

The proofs are elementary, and involve Axioms D_B and 5_B (for the first theorem), and Axioms D_B and 5_{IB} (for the second theorem). Considering Axioms 4_B and 4_{IB}, we then derive that:

$\vdash B_a\varphi \equiv B_aB_a\varphi$,
$\vdash I_a\varphi \equiv B_aI_a\varphi$.

---

8 The *canonical frame* is the frame underlying the canonical model (for details, see van der Hoek, 1993).

4. Common Belief

I now define the logic BI_1 (of belief, intention, and common belief) as an extension of BI_0.

The formal language L_1

Let us extend language L_0 by adding two more modal operators, B_E and B_*, with the following intended interpretations:

- $B_E\varphi$ means that *everybody* (i.e., each agent in A) *believes that* φ;
- $B_*\varphi$ means that φ *is a common belief* (of A).

B_E is also known as the *shared belief* operator. The language thus obtained will be denoted by L_1.

Semantics of L_1

I shall follow the most recent treatments of shared and common belief, by introducing the following relations between possible worlds:

- $\mathtt{B}_E = \bigcup_{a\in A} \mathtt{B}_a$;
- $\mathtt{B}_* = \mathtt{B}_E{}^*$, where the operator * denotes the transitive closure of a relation.[9]

The interpretation of $B_E\varphi$ and $B_*\varphi$ is then defined in the obvious way:

$$[\![B_E\varphi]\!]_M = \{w\in W\text{: } w\mathtt{B}_E w' \Rightarrow w'\in [\![\varphi]\!]_M\},$$
$$[\![B_*\varphi]\!]_M = \{w\in W\text{: } w\mathtt{B}_* w' \Rightarrow w'\in [\![\varphi]\!]_M\}.$$

The reference class of models for BI_1 is still $\mathtt{BI}$ (supplemented with the above definitions of $\mathtt{B}_E$ and $\mathtt{B}_*$). Validity of sentences of L_1 with respect to $\mathtt{BI}$ is again denoted by $\models \varphi$.

The system BI_1

I define BI_1 as the system obtained by supplementing BI_0 with the two axioms:

$$(\mathrm{E}_B) \qquad B_E\varphi \equiv \textstyle\bigwedge_{a\in A} B_a\varphi,$$
$$(\mathrm{FF}_B) \qquad B_*\varphi \supset B_E(\varphi \wedge B_*\varphi),$$

and the inference rule:

$$(\mathrm{RI}_B) \qquad \frac{\psi \supset B_E(\varphi \wedge \psi)}{\psi \supset B_*\varphi}.$$

I call FF_B the *forward fixpoint axiom* (for common belief), and RI_B the *rule of induction* (for common belief). Theorems of BI_1 are still denoted by $\vdash \varphi$.

It is easy to check that BI_1 is sound with respect to $\mathtt{BI}$: E_B is obviously true in all models of $\mathtt{BI}$, given the definition of $\mathtt{B}_E$; FF_B is true in all models such that the accessibility relation

[9] The *transitive closure* of a binary relation $\mathcal{R}$ is defined as the least transitive extension of $\mathcal{R}$.

of B_* contains the transitive closure of $\mathcal{B}_E$; and RI_B is valid if the accessibility relation of B_* is contained in the transitive closure of $\mathcal{B}_E$.

It is well known that various multi-agent epistemic logics are still sound and complete when they are augmented with E_B, FF_B, and RI_B (Halpern and Moses, 1992; see Lismont and Mongin, 1995, and Bonanno, 1996, for completeness proofs of similar systems). To my knowledge, however, a full completeness proof for such an extension of $\mathrm{KD45}_n$ is not available in the literature; in any case, no completeness proof exists for a logic of belief, common belief and intention. In the Appendix, I adapt a proof by Lismont and Mongin (1995) to a completeness proof of my logic of communication (see next section); from this, a completeness proof for BI_1 can be extracted.

5. Communication

I now extend BI_1 to the logic BI_2 (of belief, intention, common belief, and communication). The starting point to define communication is the fixpoint axiom

$$(\mathrm{F}_C) \qquad C_a\varphi \equiv B_*(\varphi \wedge I_a C_a\varphi),$$

conceptually justified in Section 2. First we have to augment our formal language.

The formal language L_2

Let us extend language L_1 by adding one more family of modal operators, $(C_a)_{a\in A}$, with the following intended interpretation:

- $C_a\varphi$ means that agent a, the *speaker*, communicates that φ to all agents in $A-\{a\}$, the *audience*.

The language thus obtained will be denoted by L_2.

Semantics of L_2

I want now to define the semantics of L_2 according to the following requirements:[10]

(i) if possible, C_a should be given a Kripkean semantics within the class of models $\mathcal{BI}$;

(ii) the semantics of C_a should enforce the validity of the fixpoint axiom F_C;

(iii) if more than one definition satisfies (i) and (ii), then the semantics of C_a should be made as weak as possible (i.e., $C_a\varphi$ should be true at the largest set of possible worlds, compatibly with (i) and (ii)).

Requirement (i) is met by defining the semantics of C_a in terms of an accessibility relation $\mathcal{C}_a$ in the usual way:

$$[\![C_a\varphi]\!]_M = \{w\in W: w\mathcal{C}_a w' \Rightarrow w'\in [\![\varphi]\!]_M\},$$

10 A similar approach could be adopted to define the semantics of common belief. I have not followed it because it would just lead to the well-known definition of the accessibility relation for common belief in terms of transitive closure.

provided we can find a relation C_a that meets the other requirements. To satisfy requirement (ii), let us define the function:

$$\gamma_a: P(W^2) \rightarrow P(W^2),$$
$$\gamma_a(R) = B_* \cup (B_* \circ I_a \circ R),$$

where $\circ$ denotes the composition of binary relations.[11]

The function γ_a maps binary relations on W onto binary relations on W. It is easy to verify that for the fixpoint axiom F_C to be valid in a class of frames, it is necessary and sufficient that the accessibility relation C_a be a fixpoint of γ_a:

$$\gamma_a(C_a) = C_a.$$

The function γ_a is monotonic; that is:

$$R \subseteq S \Rightarrow \gamma_a(R) \subseteq \gamma_a(S).$$

Given that $\langle P(W^2), \subseteq \rangle$ is a complete lattice, it follows from Tarski's fixpoint theorem (1955) that the set of fixpoints of γ_a, $fix(\gamma_a)$, is not empty, and is itself a complete lattice with respect to $\subseteq$. In general, $fix(\gamma_a)$ will contain more than one element; this means that the fixpoint axiom F_C, by itself, is not sufficient to characterize communication unambiguously, as anticipated in the Introduction.

We are now left with requirement (iii). Given that $\langle fix(\gamma_a), \subseteq \rangle$ is a complete lattice, γ_a has a least fixpoint, $\bigcap fix(\gamma_a)$, which is the smallest relation satisfying $\gamma_a(C_a) = C_a$. The three requirements for C_a are thus met if we define C_a as:

$$C_a = \bigcap fix(\gamma_a).$$

It remains now to see whether we can characterize C_a in a more constructive way. In fact, the function γ_a is upward continuous; that is, for any increasing chain

$$R_0 \subseteq R_1 \subseteq \ldots \subseteq R_n \subseteq \ldots,$$

we have

$$\gamma_a(\cup_n R_n) = \cup_n \gamma_a(R_n).$$

Therefore, the least fixpoint of γ_a is the limit of the chain:

$$R_0 = \emptyset,$$
$$R_{n+1} = \gamma_a(R_n).$$

Such a limit is given by the relation:

$$C_a = B_* \cup (B_* \circ I_a \circ B_*) \cup (B_* \circ I_a \circ B_* \circ I_a \circ B_*) \cup \ldots$$
$$= B_* \circ (1 \ \cup (I_a \circ B_*)^*),$$

---

[11] I remind the reader that the composition $\mathcal{R} \circ \mathcal{S}$ of two binary relations $\mathcal{R}$ and $\mathcal{S}$ is defined as follows:

$$x(\mathcal{R} \circ \mathcal{S})y \Leftrightarrow \exists z \,.\, x\mathcal{R}z \;\&\; z\mathcal{S}y.$$

where 1 denotes the identity relation. We now have a full Kripkean semantics for BI_2. The reference class of models for BI_2 is still $\mathcal{BI}$ (supplemented with the above definition of $\mathcal{C}_a$). I continue to denote valid sentences of L_2 by $\models \varphi$.

The system BI_2

The modal system BI_2 is defined by adding to BI_1 the forward fixpoint axiom for communication,

$$(FF_C) \quad C_a\varphi \supset B_*(\varphi \wedge I_a C_a \varphi),$$

and the rule of induction for communication,

$$(RI_C) \quad \frac{\psi \supset B_*(\varphi \wedge I_a \psi)}{\psi \supset C_a \varphi}.$$

Theorems of BI_2 will still be denoted by $\vdash \varphi$.

Again, it is easy to check that BI_2 is sound with respect to $\mathcal{BI}$: FF_C is true in all models such that the accessibility relation of C_a contains the relation $\mathcal{C}_a$ defined in the previous subsection; and RI_C is valid if the accessibility relation of C_a is contained in such relation $\mathcal{C}_a$. In the Appendix, I prove that BI_2 is also complete with respect to $\mathcal{BI}$.

As we should expect, F_C is a theorem of BI_2: the implication from left to right is FF_C, and its converse can be proved by RI_C. The repeated use of FF_C allows us to derive the following theorems:

$$\vdash C_a\varphi \supset B_*\varphi,$$
$$\vdash C_a\varphi \supset B_*I_aB_*\varphi,$$
$$\vdash C_a\varphi \supset B_*I_aB_*I_aB_*\varphi,$$
$$\vdash \ldots .$$

As one can observe, the first theorem states that the main effect the speaker wants to obtain through communication (i.e., that φ be common belief) has been achieved; the second theorem represents the achievement of a Gricean condition (i.e., that the intention to obtain the main effect also be common belief); the third theorem represents the achievement of a Strawsonian condition; and the further theorems represent the achievement of analogous conditions, for all higher orders.

In the next section, I prove further properties of communication.

6. Some properties of communication

Before proving a few facts about communication, I shall introduce some new terminology. Basically, I define terms to qualify sentences for which certain kinds of theorems can be proved in the logic. The main reason for doing so is to identify a number of concepts that play an important role in the treatment of communication.

Let us consider the theorems of the form

$$\vdash \psi \supset B_E\varphi.$$

If such a theorem can be proved for given sentences ψ and φ, we can think of ψ as describing an event that makes φ evident to all agents. In such a case, I say that ψ *displays* φ. When $\psi = \varphi$, that is when

$$\vdash \varphi \supset B_E\varphi.$$

I follow the existing practice and call φ *public*. When φ is public, it follows immediately from RI_B (with $\psi = \varphi$) that

$$\vdash \varphi \supset B_*\varphi.$$

Let us now consider the theorems of the form

$$\vdash \varphi \supset I_a\varphi.$$

If such a theorem can be proved for a given sentence φ, I say that φ is *intentional*. Interestingly, when φ is intentional we can regard it as describing an action performed by agent a: in fact, actions are basically those propositions for which being intentional is a constitutive property. Therefore, when $\varphi \supset I_a\varphi$ is a theorem I shall also say that φ *describes an action* (*by a*).

Finally, I want to define the concept of a *communicative act*. This is simply an action that implies the communication that φ, for some given sentence φ. More precisely, ψ *describes a* (φ-)*communicative act* (*by a*) if and only if:

$$\vdash \psi \supset (I_a\psi \wedge C_a\varphi).$$

An alternative characterization of communication

As defined so far, communication has a number of interesting properties. The first one, which is an immediate consequence of FF_C, is that communication implies common belief:

(C1) $\vdash C_a\varphi \supset B_*\varphi.$

This is certainly reasonable when communication is of the assertive type (and probably also in a more general context: see Section 2). Instead, the converse implication, $B_*\varphi \supset C_a\varphi$, is not a theorem of BI_2, as can be shown by a countermodel; this guarantees that communication does not merely coincide with common belief.

The second important property is that communication is intentional. From FF_C, FF_B, and E_B we derive that $\vdash C_a\varphi \supset B_aI_aC_a\varphi$, from which, by the theorem $\vdash B_aI_a\varphi \supset I_a\varphi$ (Section 3), we derive:

(C2) $\vdash C_a\varphi \supset I_aC_a\varphi.$

The converse implication, $I_aC_a\varphi \supset C_a\varphi$, is not a theorem of BI_2; hence, the intention to communicate, by itself, is not sufficient to realize communication.

A third property is that communication is public; that is:

(C3) $\vdash C_a\varphi \supset B_EC_a\varphi.$

This is a consequence of F_C, through the theorem $\vdash B_*\varphi \supset B_EB_*\varphi$. From C3 we then derive by induction that:

(C3') $\vdash C_a\varphi \supset B_*C_a\,\varphi.$

In this case, the converse implications hold:

(C4) $\vdash B_E C_a \varphi \supset C_a \varphi$,

(C4') $\vdash B_* C_a \varphi \supset C_a \varphi$.

Theorem C4 is a consequence of F_C, through the theorem $\vdash B_E B_* \varphi \supset B_* \varphi$, and C4' is a direct consequence of C4. Property C4 is very interesting, because it shows that there is nothing really "objective" in communication: if all agents individually believe that communication takes place, then communication does actually take place.

It is also possible to show that $C_a \varphi$ is the weakest statement in BI_2 that implies mutual belief that φ, and is both intentional and public. That is, if there is a statement ψ for which the following theorems can be proved:

$\vdash \psi \supset B_* \varphi$,
$\vdash \psi \supset I_a \psi$,
$\vdash \psi \supset B_E \psi$,

then it follows that:

$\vdash \psi \supset C_a \varphi$.

We can restate this condition in the form of a derived inference rule:

(C5) $$\frac{\psi \supset B_* \varphi,\ \psi \supset I_a \psi,\ \psi \supset B_E \psi}{\psi \supset C_a \varphi}.$$

To prove the validity of C5, it is sufficient to remark that if ψ satisfies the assumption of C5, then it also satisfies the assumption of RI_C (apply RI_B to $\psi \supset B_E \psi$, and the monotonicity property of B_* to $\psi \supset I_a \psi$). We have now an interesting alternative characterization of communication: in BI_2, $C_a \varphi$ is the weakest intentional and public statement that implies the common belief that φ.

The fact that communication is intentional suggests that the statement $C_a \varphi$ describes some kind of action. However, a sentence of the form $C_a \varphi$ need not by itself describe a *basic* action.[12] In general, to communicate that φ one has to perform some lower-level action ψ that realizes $C_a \varphi$; typically, ψ will be some action conventionally expressing a meaning, like raising one's hand, or ringing a bell, or uttering a sentence in a natural language. We now face a fundamental problem: under what conditions does a lower-level action realize a communicative act?

---

[12] A *basic action* is an action that can be performed immediately, with no need of a lower-level action to realize it. Examples of basic actions are raising one's arm and emitting a sequence of vocal sounds.

One such a set of conditions is shown by rule C5: an intentional statement describes a φ-communicative act if it is public and it implies the common belief that φ. However, we can prove the validity of a more interesting inference rule, which I shall call the *rule of communication*:

$$\text{(RC)} \qquad \frac{\psi \supset I_a\psi,\ \psi \supset B_E(\varphi \wedge \psi)}{\psi \supset C_a\varphi}.$$

The derivation of RC is immediate: From $\psi \supset B_E(\varphi \wedge \psi)$ we derive through RI_C that $\psi \supset B_*(\varphi \wedge \psi)$. From this and $\psi \supset I_a\psi$, we then derive that $\psi \supset B_*(\varphi \wedge I_a\psi)$, and from this we derive by RI_C that $\psi \supset C_a\varphi$.

The rule of communication can be paraphrased as follows: a sufficient condition for an intentional statement ψ to describe a φ-communicative act is that ψ is public and displays φ. Now we can go back to the problem I have put forward in the Introduction: how can agents infer communication from individual beliefs? The answer comes by combining RC with C4 (exploiting the monotonicity of B_E); we thus derive the inference rule:

$$\text{(RC')} \qquad \frac{\psi \supset I_a\psi,\ \psi \supset B_E(\varphi \wedge \psi)}{B_E\psi \supset C_a\varphi},$$

which says that if an intentional statement ψ is public and displays φ, for φ to be communicated it is sufficient that the speaker and all agents in the audience individually believe that ψ is performed. This rule gives a possible solution to the problem of inferring communication from individual beliefs.

7. Discussion

Within an appropriate propositional modal logic, I have defined the proposition '*a communicates that* φ' as the largest set of possible worlds at which it is common belief that φ, and that a intends to communicate that φ. On the basis of such a definition, I have given sufficient conditions for an action to be communicative, and for communication to take place.

It seems that the two rules of communication presented in Section 6, RC and RC', bring us closer to understanding how one can communicate something by performing a lower-level action. For example, let us assume that intentionally raising one's arm in the appropriate context (ψ) expresses the request to ask a question (φ). Let us further suppose that in the previously mentioned "appropriate context" we can safely assume that raising one's arm is intentional and public. If Alice intentionally raises her arm in such a context, we can derive by RC that Alice communicates to her audience the request to ask a question.

It is important to interpret this result correctly. The rules of communication do *not* say that, for any model M, a communicates that φ by performing ψ at any world of M where $\psi \supset I_a\psi$ and $\psi \supset B_E(\varphi \wedge \psi)$ are true (in fact, this statement is false, as it can be shown through a counterexample). Rather, they prove a weaker assertion: namely, that a communicates that φ by performing ψ at *all* worlds of *all* models, if $\psi \supset I_a\psi$ and $\psi \supset B_E(\varphi \wedge \psi)$ are true at *all* worlds of *all* models. Going back to the previous example, this means that there must be no exception to the fact that raising one's arm in the appropriate context is intentional, is public, and displays the request to ask a question.

These considerations show that the rule of communication requires a communicative act to be construed as a very abstract act type – so abstract that it might seem completely unrealistic. This degree of abstraction is however common to many formal theories of human action. Typically, such theories are built on a formal language containing a class of expressions that describe actions; for the sake of our present discussion, let such expressions have the form $do(a,e)$, where a names the agent and e denotes the action performed. The theory will typically then contain an axiom like

$$do(a,e) \supset I_a\, do(a,e),$$

stating that actions are intentional. But this is precisely my requirement on statements that describe actions. Similar considerations apply to the other assumptions I have exploited to provide conditions for communication. All such conditions are very strong, because they are expressed in terms of theoremhood, and not of truth at a specific world.

This limitation reflects a very basic fact: no concrete data about a specific act token can be sufficient to logically prove that a communicative act has been performed. Let us go back again to the arm raising example. In a concrete situation, if Alice's arm goes up, her audience cannot establish beyond any doubt that Alice raised her arm intentionally, and that she did so in order to mean "I'd like to ask a question". Therefore, that Alice performed a communicative act cannot be established as a logical theorem. As a consequence of such considerations, it would seem as if the recognition of a communicative act could only be based on some kind of nonmonotonic reasoning, allowing the audience to jump to conclusions that would not be fully licensed from a strictly logical point of view (see also Perrault , 1990, and Airenti *et al.*, 1993). For example, agents might exploit rules of the following type:

- raising one's arm is intentional *by default*;
- raising one's arm in the appropriate context (e.g., when every agent can observe every other agent) is public *by default*;
- raising one's arm in the appropriate context (e.g., during a lecture) *by default* displays the request to ask a question.

To find out which kind of nonmonotonic logic is best suited for this type of problems is an interesting topic for future investigation.

Another interesting research issue is the connection between communication and belief revision. In this article I have completely neglected the temporal dimension, but of course the main reason for a speaker to communicate something is to *change* the mental state of his or her audience. I expect that taking into account the intentional aspects of human interaction will make it possible to provide a more realistic model of the process by which a group of agents can reach a reasonably shared view of the world, which in turn is a necessary precondition for various types of cooperative and competitive interaction.

Appendix: Proof of Completeness for BI_2

To prove the completeness of BI_2, I adapt a similar proof by Lismont and Mongin (1995; henceforward LM).

Preliminary elements

Given a propositional multi-modal language L based on a set P of propositional constants, and a sentence $\varphi \in L$, the *depth*, $dp(\varphi)$, of φ is defined as:

- $dp(p) = 0$, for all $p \in P$;
- $dp(\sim\varphi) = dp(\varphi)$;
- $dp(\varphi \# \psi) = max(dp(\varphi),dp(\psi))$, where # is any binary Boolean connective;
- $dp(O\varphi) = dp(\varphi)+1$, where O is any modal operator.

For any sentence $\chi \in L$, the set $L[\chi]$ is defined as the set of all sentences $\varphi \in L$, built from the propositional constants occurring in χ, and such that $dp(\varphi) \leq dp(\chi)$.

Given a multi-modal logical system Σ based on L, a subset $\Gamma \subseteq L[\chi]$ is said to be $\Sigma[\chi]$-*maximal consistent* if and only if: Γ is Σ-consistent, and $\Gamma \cup \{\varphi\}$ is Σ-inconsistent for all $\varphi \in L[\chi]$ such that $\varphi \notin \Gamma$. The set of all $\Sigma[\chi]$-maximal consistent sets will be denoted by $MaxCon_{\Sigma[\chi]}$, and the set of all $\Sigma[\chi]$-maximal consistent sets containing a given sentence $\varphi \in L[\chi]$ will be denoted by $[\varphi]_{\Sigma[\chi]}$. Note that if $\Gamma \in MaxCon_{\Sigma[\chi]}$, then:

- $\varphi \in \Gamma$ if and only if $\Gamma \in [\varphi]_{\Sigma[\chi]}$;
- $\varphi \in \Gamma$ if and only if $\varphi \in L[\chi]$ and $\Gamma \vdash \varphi$.

Some properties of $\Sigma[\chi]$-maximal consistent sets

The $\Sigma[\chi]$-maximal consistent sets are related to the classical maximal consistent sets in the following way: $MaxCon_{\Sigma[\chi]}$ is the set of all intersections $\Gamma' \cap L[\chi]$, where Γ' ranges over the set of all Σ-maximal consistent subsets of L. An important consequence is:

Lemma 1. *For any sentence χ:*

$[\chi]_{\Sigma[\chi]} = MaxCon_{\Sigma[\chi]} \Rightarrow \vdash \chi$.

Proof. If $[\chi]_{\Sigma[\chi]} = MaxCon_{\Sigma[\chi]}$, then $\chi \in \Gamma$ for all $\Sigma[\chi]$-maximal consistent sets Γ. Hence $\chi \in \Gamma' \cap L[\chi]$ for all Σ-maximal consistent sets Γ', which implies $\vdash \chi$. •

The following is Lemma 2 by LM:

Lemma 2. *For any sentence χ:*

(1) *The set $MaxCon_{\Sigma[\chi]}$ is finite.*

(2) *For any set $X \subseteq MaxCon_{\Sigma[\chi]}$, there is a sentence $\varphi \in L[\chi]$ such that $[\varphi]_{\Sigma[\chi]} = X$.*

(3) *For all φ, $\psi \in L[\chi]$, $[\varphi]_{\Sigma[\chi]} \subseteq [\psi]_{\Sigma[\chi]} \Leftrightarrow \vdash \varphi \supset \psi$.* •

$\Sigma[\chi]$-canonical models

Given a sentence $\chi \in L$, a $\Sigma[\chi]$-*canonical model* is a model M_χ built in the following way:

- the set of possible worlds of M_χ is $MaxCon_{\Sigma[\chi]}$ (possible worlds therefore coincide with $\Sigma[\chi]$-maximal consistent sets);
- the valuation v of M_χ is defined by $v(p) = [p]_{\Sigma[\chi]}$.

This far, the accessibility relations of M_χ are not specified.

Plan of the completeness proof

Given a language L, a reference class of models $\mathtt{M}$ for L, and a logical system Σ based on L, suppose that for each sentence $\chi \in L$ we are able to build a $\Sigma[\chi]$-canonical model M_χ such that:

(i) M_χ belongs to $\mathtt{M}$;

(ii) for all sentences $\varphi \in L[\chi]$, $[\![\varphi]\!]_{M_\chi} = [\varphi]_{\Sigma[\chi]}$.

It follows that Σ is complete (with respect to $\mathtt{M}$). In fact, for a given sentence $\chi \in L$, let us assume that $\models \chi$; then, by (i), $M_\chi \models \chi$. This implies that $[\![\chi]\!]_{M_\chi} = MaxCon_{\Sigma[\chi]}$, and therefore, by (ii), $[\chi]_{\Sigma[\chi]} = MaxCon_{\Sigma[\chi]}$, which in turn implies $\vdash \chi$ by Lemma 1.

To prove the completeness of BI_2, for each sentence $\chi \in L$ I first define a $\mathrm{BI}_2[\chi]$-canonical model M_χ, so that it belongs to $\mathtt{BI}$. Then I prove (ii), the so-called "fundamental lemma".

The $\mathrm{BI}_2[\chi]$*-canonical model*

Given a sentence χ of L_2, we already know how to define the set of possible worlds and the valuation of M_χ. As for the accessibility relations, let me first introduce the relations $\mathtt{R}_a$ and $\mathtt{S}_a$ in the following way:

- $\Gamma \mathtt{R}_a \Delta \Leftrightarrow \{\varphi \in L_2[\chi]: \Gamma \vdash B_a\varphi\} \subseteq \Delta$;
- $\Gamma \mathtt{S}_a \Delta \Leftrightarrow \{\varphi \in L_2[\chi]: \Gamma \vdash I_a\varphi\} \subseteq \Delta$.

I now define the accessibility relations $\mathtt{B}_a$ and $\mathtt{I}_a$ of M_χ as follows[13]:

$$\Gamma \mathtt{B}_a \Delta \Leftrightarrow \exists n \geq 0 \,.\, \Gamma (\mathtt{R}_a')^n \circ \mathtt{R}_a \Delta,$$

$$\Gamma \mathtt{I}_a \Delta \Leftrightarrow \exists n \geq 0 \,.\, \Gamma (\mathtt{R}_a')^n \circ \mathtt{S}_a \Delta,$$

where by $\mathtt{R}_a'$ I denote the union of $\mathtt{R}_a$ and of its converse $\mathtt{R}_a^{-1}$ (the *converse* $\mathtt{R}^{-1}$ of a binary relation $\mathtt{R}$ is defined by: $x\mathtt{R}^{-1}y$ if and only if $y\mathtt{R}x$).

Let us check that the model M_χ belongs to $\mathtt{BI}$.

[13] It is possible to prove that $\mathcal{B}_a$ is the *transitive and Euclidean closure of* $\mathcal{R}_a$, that is, the least transitive and Euclidean extension of $\mathcal{R}_a$. However, this fact is not relevant to the current completness proof. Analogously, $\mathcal{I}_a$ is the *transitive and Euclidean closure of* $\mathcal{S}_a$ *over* $\mathcal{R}_a$, that is, the least extension of $\mathcal{S}_a$ which is transitive over $\mathcal{R}_a$ and Euclidean over $\mathcal{R}_a$.

Lemma 3. *M_χ belongs to $\mathtt{BI}$.*

Proof. I have only to check that the accessibility relations have the required properties:

(1) *B_a is serial*, because it extends R_a, which is a serial relation. To prove that R_a is serial, I have to show that for all $\Gamma \in MaxCon_{\Sigma[\chi]}$ there is a $\Delta \in MaxCon_{\Sigma[\chi]}$ such that

$$\{\varphi \in L_2[\chi]: \Gamma \vdash B_a\varphi\} \subseteq \Delta.$$

The set on the left is not empty (thanks to RN_B) and consistent (thanks to D_B). Therefore it can be extended to a $BI_2[\chi]$-maximal consistent set Δ.

(2) *B_a is transitive*, that is, $B_a \circ B_a \subseteq B_a$:

$$\begin{aligned} \Gamma B_a \circ B_a \Delta &\Leftrightarrow \exists i \geq 0\ \exists k \geq 0\ .\ \Gamma (R_a')^i \circ R_a \circ (R_a')^k \circ R_a \Delta \\ &\Rightarrow \exists n \geq 0\ .\ \Gamma (R_a')^n \circ R_a \Delta \Leftrightarrow \Gamma B_a \Delta \end{aligned}$$

(3) *B_a is Euclidean*, that is, $B_a^{-1} \circ B_a \subseteq B_a$:

$$\begin{aligned} \Gamma B_a^{-1} \circ B_a \Delta &\Leftrightarrow \exists i \geq 0\ \exists k \geq 0\ .\ \Gamma R_a^{-1} \circ (R_a')^i \circ (R_a')^k \circ R_a \Delta \\ &\Rightarrow \exists n \geq 0\ .\ \Gamma (R_a')^n \circ R_a \Delta \Leftrightarrow \Gamma B_a \Delta \end{aligned}$$

(4) *I_a is serial*, because it includes S_a, which is a serial relation. That S_a is serial is proved by an argument similar to case (1) above, based on RN_I and D_I.

(5) *I_a is transitive over B_a*, that is, $B_a \circ I_a \subseteq I_a$: the proof is analogous to case (2) above.

(6) *I_a is Euclidean over B_a*, that is, $B_a^{-1} \circ I_a \subseteq I_a$: analogous to case (3) above. •

The completeness of BI_2

To prove the completeness of BI_2, I have now to prove the fundamental lemma: for all sentences $\varphi \in L_2[\chi]$, $[\![\varphi]\!]_{M_\chi} = [\varphi]_{BI_2[\chi]}$. From now on, the indexes ‘$_{M_\chi}$’ and ‘$_{BI_2[\chi]}$’ are dropped to improve readability.

Before I proceed further, let me introduce some helpful notation (following LM). Let L be a multi-modal language with Kripkean semantics. For each modal operator O, with accessibility relation R, I define a function r: $P(W) \to P(W)$ associated to R as:

$$r(X) = \{w \in W: wRw' \Rightarrow w' \in X\}.$$

It follows that:

$$[\![O\varphi]\!]_M = r([\![\varphi]\!]_M).$$

It is easy to prove that r is downward continuous; that is, for any decreasing chain

$$X_0 \supseteq X_1 \supseteq \ldots \supseteq X_n \supseteq \ldots,$$

we have

$$r(\cap_n X_n) = \cap_n r(X_n).$$

This implies that r is monotone:

$$X \subseteq Y \Rightarrow r(X) \subseteq r(Y).$$

Moreover, if R and S are two accessibility relations on W, and r and s are the two functions respectively associated to R and S, we have:

$r(s(X)) = \{w \in W: wR \circ Sw' \Rightarrow w' \in X\}$.

The functions associated to B_a, I_a, B_E, B_*, and C_a will be respectively denoted by b_a, i_a, b_E, b_*, and c_a. We have the following lemma.

Lemma 4. *In all models of L_2, the following equations hold*:

(1) $b_E(X) = \cap_{a \in A} b_a(X)$;

(2) $b_*(X) = \cup\{Y: Y \subseteq b_E(X \cap Y)\}$;

(3) $c_a(X) = \cup\{Y: Y \subseteq b_*(X \cap i_a(Y))\}$.

Proof.

(1) The equation follows easily from the definition of B_E.

(2) From the definition of b_* and B_* and the properties of b_E it is easy to derive that

$$b_*(X) = \cap_{n>0} b_E{}^n(X).$$

From the monotonicity of b_E we also derive that

$$Y \subseteq \cap_{n>0} b_E{}^n(X) \Leftrightarrow Y \subseteq b_E(X \cap Y).$$

Hence the conclusion.

(3) Analogous to case (2) above, from the definitions of c_a and C_a and the properties of b_* and i_a. •

We are now ready to prove some important properties of M_χ.

Lemma 5. *For all $\varphi \in L_2[\chi]$, such that $dp(\varphi) < dp(\chi)$, the following equations hold*:

(1) $[B_a\varphi] = \{\Gamma \in MaxCon_{\Sigma[\chi]}: \Gamma R_a \Delta \Rightarrow \Delta \in [\varphi]\}$;

(2) $[I_a\varphi] = \{\Gamma \in MaxCon_{\Sigma[\chi]}: \Gamma S_a \Delta \Rightarrow \Delta \in [\varphi]\}$.

(3) $[B_a\varphi] = b_a([\varphi])$;

(4) $[I_a\varphi] = i_a([\varphi])$;

(5) $[B_E\varphi] = b_E([\varphi])$;

(6) $[B_*\varphi] = b_*([\varphi])$;

(7) $[C_a\varphi] = c_a([\varphi])$.

Proof.

(1, 2) See Lemma 6(2) by LM.

(3) The proof is similar to the one provided by LM (p. 146). Taking into account case (1) above and the definition of $b_a([\varphi])$, I have to prove that:

$$(\Gamma R_a \Delta \Rightarrow \Delta \in [\varphi]) \Leftrightarrow (\Gamma B_a \Delta \Rightarrow \Delta \in [\varphi]).$$

Given that $R_a \subseteq B_a$, the implication from right to left is obvious. To show that the converse implication also holds, I prove that, for any $n \geq 0$,

(∗) $(\Gamma \mathcal{R}_a \Delta \Rightarrow \Delta \in [\varphi])$ & $\Gamma (\mathcal{R}_a')^n \circ \mathcal{R}_a \Delta \Rightarrow \Delta \in [\varphi]$,

To do so, let me first consider two immediately subsequent worlds, Γ' and Γ'', on a $\mathcal{R}_a'$-path from Γ to Δ; then, either $\Gamma' \mathcal{R}_a \Gamma''$ or $\Gamma'' \mathcal{R}_a \Gamma'$. I show that in both cases the sentence $B_a\varphi$ can be "pushed" from Γ' to Γ''; that is, if $B_a\varphi \in \Gamma'$, then $B_a\varphi \in \Gamma''$. In fact, assume that $B_a\varphi \in \Gamma'$, and consider the case $\Gamma' \mathcal{R}_a \Gamma''$. By 4_B, we have $\Gamma' \vdash B_a B_a \varphi$, and therefore $B_a\varphi \in \Gamma''$. Alternatively, assume that $B_a\varphi \in \Gamma'$, and consider the case $\Gamma'' \mathcal{R}_a \Gamma'$. If $B_a\varphi \notin \Gamma''$, by maximality we have $\sim B_a\varphi \in \Gamma''$; then, by 5_B,,
$\Gamma'' \vdash B_a \sim B_a \varphi$, and hence $\sim B_a \varphi \in \Gamma'$, contrary to the assumption.

Now, assume the antecedent of (∗). From case (1) we have $B_a\varphi \in \Gamma$. By the previous argument, we have that $B_a\varphi$ can be pushed along the $\mathcal{R}_a'$-path, until we reach a word Δ' such that $\Delta' \mathcal{R}_a \Delta$ and $B_a\varphi \in \Delta'$. The conclusion follows immediately.

(4) The proof is analogous to that of case (3). Taking into account case (2) above and the definition of $i_a([\varphi])$, I have to show that:

$$(\Gamma \mathcal{S}_a \Delta \Rightarrow \Delta \in [\varphi]) \Leftrightarrow (\Gamma \mathcal{I}_a \Delta \Rightarrow \Delta \in [\varphi]).$$

Given that $\mathcal{S}_a \subseteq \mathcal{I}_a$, the implication from right to left is obvious. To show that the converse also holds, I prove that, for $n \geq 0$,

$$(\Gamma \mathcal{S}_a \Delta \Rightarrow \Delta \in [\varphi]) \ \& \ \Gamma (\mathcal{R}_a')^n \circ \mathcal{S}_a \Delta \Rightarrow \Delta \in [\varphi].$$

Assume the antecedent. From case (2) we have $I_a\varphi \in \Gamma$ and, by 4_{IB}, $\Gamma \vdash B_a I_a \varphi$. An argument similar to that of case (1), based on 4_{IB} and 5_{IB}, shows that we can push the sentence $I_a\varphi$ until we reach Δ' such that $I_a\varphi \in \Delta'$ and $\Delta' \mathcal{S}_a \Delta$, from which we derive the conclusion.

(5) By Axiom E_B we have that $[B_E\varphi] = \bigcap_{a \in A}[B_a\varphi]$. The conclusion follows from case (3) above and Lemma 4(1).

(6) Let me first show that for all $\Gamma \in MaxCon_{\Sigma[\chi]}$:

(∗∗) $B_*\varphi \in \Gamma \Leftrightarrow \exists \psi \in L_2[\chi] \,.\, \psi \in \Gamma \ \& \ [\psi] \subseteq b_E([\varphi] \cap [\psi])$.

Left to right. Let $\psi = B_*\varphi$; then obviously $\psi \in L_2[\chi]$ and $\psi \in \Gamma$. Moreover, by FF_B, $\vdash \psi \supset B_E(\varphi \wedge \psi)$, which implies $[\psi] \subseteq b_E([\varphi] \cap [\psi])$ by Lemma 2(3) and case (5).

Right to left. From $[\psi] \subseteq b_E([\varphi] \cap [\psi])$, by Lemma 2(3) and case (5) above, we have $\vdash \psi \supset B_E(\varphi \wedge \psi)$. Then, by RI_B, we derive $\vdash \psi \supset B_*\varphi$, and hence $B_*\varphi \in \Gamma$, because $\psi \in \Gamma$ and $B_*\varphi \in L_2[\chi]$.

From (∗∗), we conclude that

$$[B_*\varphi] = \bigcup\{[\psi]: \psi \in L_2[\chi] \ \& \ [\psi] \subseteq b_E([\varphi] \cap [\psi])\},$$

from which, through Lemmas 2(2) and 4(2), we derive

$$[B_*\varphi] = \bigcup\{Y: Y \subseteq b_E([\varphi] \cap Y)\} = b_*([\varphi]).$$

(7) The proof is analogous to that of case (6) above, and is based on FF_C and on RI_C; besides Lemmas 2(2) and 2(3), it exploits Lemma 4(3) and the result of case (6). •

I can now prove the:

Fundamental Lemma: $[\![\varphi]\!] = [\varphi]$.

Proof. The proof is by induction on the structure of sentences; throughout the proof, $\Gamma,\Delta \in MaxCon_{\Sigma[\chi]}$.

(1)	$[\![p]\!]$	$= v(p) = [p]$, for all $p \in P$	by the definition of v in M_χ;
(2)	$[\![\sim\varphi]\!]$	$= MaxCon_{\Sigma[\chi]} - [\![\varphi]\!]$	by definition,
		$= MaxCon_{\Sigma[\chi]} - [\varphi]$	by the induction hypothesis,
		$= [\sim\varphi]$	by maximality;
(3)	$[\![\varphi \# \psi]\!]$	$= [\varphi \# \psi]$, for each connective #	the proof is analogous to case (2);
(4)	$[\![B_a\varphi]\!]$	$= b_a([\![\varphi]\!])$	by definition,
		$= b_a([\varphi])$	by the induction hypothesis,
		$= [B_a\varphi]$	by Lemma 5(3);
(5)	$[\![I_a\varphi]\!]$	$= [I_a\varphi]$	analogous to case (4), by Lemma 5(4);
(6)	$[\![B_E\varphi]\!]$	$= [B_E\varphi]$	analogous to case (4), by Lemma 5(5);
(7)	$[\![B_*\varphi]\!]$	$= [B_*\varphi]$	analogous to case (4), by Lemma 5(6);
(8)	$[\![C_a\varphi]\!]$	$= [C_a\varphi]$	analogous to case (4), by Lemma 5(7). •

References

G. Airenti, B.G. Bara and M. Colombetti. Conversation and behavior games in the pragmatics of dialogue, *Cognitive Science* 17 (1993) 197–256.

J.A. Austin. *How to do things with words* (Oxford University Press, Oxford, 1962).

G. Bonanno. On the logic of common belief, *Mathematical Logic Quarterly* 42 (1996) 305–311.

B.F. Chellas. *Modal logic* (Cambridge University Press, Cambridge, 1980).

P.R. Cohen and H.J. Levesque. Rational interaction as a basis for communication. In: P.R. Cohen, J. Morgan and M.E. Pollack, eds., *Intentions in communication* (MIT Press, Cambridge, MA, 1990) pp. 221–255.

H.P. Grice. Meaning, *Philosophical Review* 67 (1957) 377–388.

J.Y. Halpern and Y. Moses. A guide to completeness and complexity for modal logics of knowledge and belief, *Artificial Intelligence* 54 (1992) 319–379.

K. Konolige and M.E. Pollack. A representational theory of intention, *Proceedings of the AAAI Spring Symposium on Reasoning about Mental States* (AAAI Press, Menlo Park, CA, 1993) pp. 178–188.

D.K. Lewis. *Convention: A philosophical study* (Harvard University Press, Cambridge, MA, 1969).

L. Lismont and P. Mongin. Belief closure: A semantics of common knowledge for modal propositional logic, *Mathematical Social Sciences* 30 (1995) 127–153.

C.R. Perrault. An application of default logic to speech act theory. In: P.R. Cohen, J. Morgan and M.E. Pollack,eds., *Intentions in communication* (MIT Press, Cambridge, MA, 1990) pp. 161–193.

S. Schiffer. *Meaning* (Oxford University Press, Oxford, 1972).

J.R. Searle. *Speech acts* (Cambridge University Press, Cambridge, 1969).

C. Shannon and W. Weaver. *The mathematical theory of communication* (University of Illinois Press, Urbana, IL, 1949).

D. Sperber and D. Wilson. *Relevance* (Basil Blackwell, Oxford, 1986).

P. Strawson. Intention and convention in speech acts, *Philosophical Review* 73 (1964) 439–460.

A. Tarski. A lattice-theoretical fixpoint theorem and its applications, *Pacific Journal of Mathematics* 5 (1955) 285–309.

W. van der Hoek. Systems of knowledge and belief, *Logic and Computation* 3 (1993) 173–195.