Group Knowledge Isn't Always Distributed *
(Neither Is It Always Implicit)

W. van der Hoek B. van Linder J.-J. Ch. Meyer
Utrecht University – Department of Computer Science
P.O. Box 80.089 – 3508 TB Utrecht – The Netherlands

Abstract

In this paper we study the notion of group knowledge in a modal epistemic context. Starting with the standard definition of this kind of knowledge on Kripke models, we show that it may behave quite counter-intuitively. Firstly, using a strong notion of derivability, we show that group knowledge in a state can always, but trivially be derived from each of the agents' individual knowledge. In that sense, group knowledge is not really *implicit*, but rather *explicit* knowledge of the group. Thus, a weaker notion of derivability seems to be more adequate. However, adopting this more 'local view', we argue that group knowledge need not be *distributed* over (the members of) the group: we give an example in which (the traditional concept of) group knowledge is stronger than what can be derived from the individual agents' knowledge. We then propose two additional properties on Kripke models: we show that together they are sufficient to guarantee 'distributivity', while, when leaving one out, one may construct models that do not fulfill this principle.

1 Introduction

In the field of AI and computer science, the modal system **S5** is a well-accepted and by now familiar logic to model the logic of knowledge (cf. [2, 7]). Since the discovery of **S5**'s suitability for epistemic logic, many extensions and adaptions of this logic have been proposed. In this paper we look into one of these extensions, namely a system for knowledge of m agents containing a modality for *'group knowledge'*. Intuitively this 'group knowledge' (let us write GK for it) is not the knowledge of each of the agents in the group, but the knowledge that would result if the agents could somehow 'combine' their knowledge. The intuition behind this notion is best illustrated by an example. Let the formula φ denote the proposition that $P \neq NP$. Assume that three computer scientists are working on a proof of this proposition. Suppose that φ follows from three lemmas: ψ_1, ψ_2 and ψ_3. Assume that scientist 1 has proved ψ_1 and therefore knows ψ_1. Analogously for agents 2 and 3 with respect to ψ_2 and $\psi_3 = (\psi_1 \wedge \psi_2) \to \varphi$. If these computer scientists would be able to contact each other at a conference, thereby combining their knowledge, they would be able to conclude φ. This example also illustrates the relevance of *communication* with respect to this kind of group knowledge: the scientists should somehow transfer their knowledge through communication in order to make the underlying implicit knowledge explicit.

In this paper, we try to make the underlying notions that together constitute GK explicit: What does it mean for a group of, say m agents, to *combine* their knowledge? We start by giving and explaining a clear semantical definition of group knowledge, as it was given by Halpern and Moses in [2]. In order to make some of our points, we distinguish between a *global* and a *local* notion of (deductive and semantic) consequence. Then we argue that the

*This work was partially supported by ESPRIT III BRWG No. 8319 (ModelAge) and ESPRIT III BRA No. 6156 (DRUMS II).

defined notion of group knowledge may show some counter-intuitive behavior. For instance, we show in Section 3 that, at the global level, group knowledge does not add deductive power to the system: group knowledge is not always a true refinement of individual knowledge. Then, in Section 4, we try to formalize what it means to combine the knowledge of the members of a group. We give one principle (called the principle of distributivity: it says that GK that is derived from a set of premises, can always be derived from a conjunction of m formulae, each known by one of the agents) that is trivially fulfilled in our set-up. We also study a special case of this definition (called the principle of full communication), and show that this property is not fulfilled when using standard definitions in standard Kripke models. In Section 5 we induce two additional properties on Kripke models; we show that they are sufficient to guarantee full communication: they are also necessary in the sense that one can construct models that do not obey one of the additional properties and that at the same time do not verify the principle under consideration. In Section 6 we round off.

2 Knowledge and Group Knowledge

Halpern and Moses introduced an operator to model group knowledge ([2]). Initially this knowledge in a group was referred to as 'implicit knowledge' and indicated with a modal operator I. Since in systems for knowledge and belief, the phrase 'implicit' already had obtained its own connotation (cf. [5]), later on, the term 'distributed knowledge' (with the operator D) became the preferred name for the group knowledge we want to consider here (cf. [3]). Since we do not want to commit ourselves to any fixed terminology we use the operator G to model the 'group knowledge'. From the point of view of communicating agents, G-knowledge may be seen as the knowledge being obtained if the agents were fully able to communicate with each other. Actually, instead of being able to communicate *with each other*, one may also adopt the idea that the G-knowledge is just the knowledge of one distinct agent, to whom all the agents communicate their knowledge (this agent was called the 'wise man' in [2]; a system to model such communication was proposed in [8]). We will refer to this reading of G (i.e., in a 'send and receiving context') as 'a receiving-agent's knowledge'.

We start by defining the language that we use.

Definition 2.1 Let Π be a non-empty set of propositional variables, and $m \in \mathbb{N}$ be given. The language $\mathcal{L}$ is the smallest superset of Π such that:

$$\text{if } \varphi, \psi \in \mathcal{L} \text{ then } \neg\varphi, (\varphi \wedge \psi), K_i\varphi, G\varphi \in \mathcal{L} \ (i \leq m)$$

The familiar connectives $\vee, \rightarrow, \leftrightarrow$ are introduced by definitional abbreviation in the usual way; $\top \stackrel{\text{def}}{=} p \vee \neg p$ for some $p \in \Pi$ and $\bot \stackrel{\text{def}}{=} \neg\top$.

The intended meaning of $K_i\varphi$ is 'agent i knows φ' and $G\varphi$ means 'φ is group knowledge of the m agents.' We assume the following 'standard' inference system $\mathbf{S5}_m(G)$ for the multi-agent **S5** logic incorporating the operator G.

Definition 2.2 The logic $\mathbf{S5}_m(G)$ has the following axioms:

$A1$ any axiomatization for propositional logic

$A2$ $(K_i\varphi \wedge K_i(\varphi \rightarrow \psi)) \rightarrow K_i\psi$

$A3$ $K_i\varphi \rightarrow \varphi$

$A4$ $K_i\varphi \rightarrow K_iK_i\varphi$

$A5$ $\neg K_i\varphi \rightarrow K_i\neg K_i\varphi$

$A6$ $K_i\varphi \rightarrow G\varphi$

$A7$ $(G\varphi \wedge G(\varphi \rightarrow \psi)) \rightarrow G\psi$

$A8$ $G\varphi \rightarrow \varphi$

$A9$ $G\varphi \rightarrow GG\varphi$

$A10$ $\neg G\varphi \rightarrow G\neg G\varphi$

On top of that, we have the following derivation rules:

$R1$ $\vdash \varphi, \vdash \varphi \rightarrow \psi \Rightarrow \vdash \psi$

$R2$ $\vdash \varphi \Rightarrow \vdash K_i\varphi$, for all $i \leq m$

In words, we assume a logical system $(A1, R1)$ for *rational* agents. Individual knowledge, i.e., the knowledge of one agent, is moreover supposed to be *veridical* $(A3)$. The agents are assumed to be *fully introspective*: they are supposed to have *positive* $(A4)$ as well as *negative* $(A5)$ introspection. In the receiving-agent's reading, the axiom $A6$ may be understood as the 'communication axiom'; what is known to some member of the group is also known to the receiving agent. The other axioms express that the receiver has the same reasoning and introspection properties as the other agents of the group. In the group knowledge reading, $A6$ declares what the members of the group are, the other axioms enforce this group knowledge to obey the same properties that are ascribed to the individual agents.

The derivability relation $\vdash_{\mathbf{S5}_m(G)}$, or $\vdash$ for short, is defined in the usual way. That is, a formula φ is said to be *provable*, denoted $\vdash \varphi$, if φ is an instance of one of the axioms or if φ follows from provable formulae by one of the inference rules $R1$ and $R2$. We define two variants of *provability from premises*: one in which necessitation on premises is allowed and one in which it is not.

Definition 2.3 Let ψ be some formula, and let Φ be a set of formulae. Using the relation $\vdash$ of provability within the system $\mathbf{S5}_m(G)$ we define the following two relations $\vdash^+$ and $\vdash^- \subseteq 2^{\mathcal{L}} \times \mathcal{L}$:

$\Phi \vdash^+ \psi$ $(\Phi \vdash^- \psi)$ $\Leftrightarrow \exists \varphi_1, \ldots, \varphi_n$ with $\varphi_n = \psi$, and such that for all $1 \leq i \leq n$:

- either $\varphi_i \in \Phi$
- or there are $j, k < i$ with $\varphi_j = \varphi_k \rightarrow \varphi_i$
- or φ_i is an $\mathbf{S5}_m(G)$ axiom
- or $\varphi_i = K_k\sigma$ where σ is such that

$$\begin{cases} \sigma = \varphi_j \text{ with } j < i. & \text{in case of } \vdash^+ \\ \vdash \sigma & \text{in case of } \vdash^- \end{cases}$$

So, the relation $\vdash^+$ is more liberal than $\vdash^-$ in the sense that $\vdash^+$ allows for necessitation on premises, where $\vdash^-$ only applies necessitation to $\mathbf{S5}_m(G)$-theorems. From a modal logic point of view, one establishes:

$$\vdash (\varphi_1 \wedge \ldots \wedge \varphi_n) \rightarrow \psi \;\Leftrightarrow\; \varphi_1, \ldots, \varphi_n \vdash^- \psi \;\Rightarrow\; \varphi_1, \ldots, \varphi_n \vdash^+ \psi$$

Also, one can prove the following connection between the two notions of derivability from premises ([6]):

$$K_1\varphi_1, \ldots, K_m\varphi_1, \ldots, K_1\varphi_n, \ldots, K_m\varphi_n \vdash^- \psi \;\Leftrightarrow\; \varphi_1, \ldots, \varphi_n \vdash^+ \psi$$

From an epistemic point of view, when using $\vdash^-$, we have to view a set of premises Φ as a set of additional *given* (i.e., true) formulae, whereas in the case of $\vdash^+$, Φ is a set of *known* formulae.

Definition 2.4 A Kripke model $\mathcal{M}$ is a tuple $\mathcal{M} = \langle \mathcal{S}, \pi, R_1, \ldots, R_m \rangle$ where

1. $\mathcal{S}$ is a non-empty set of states,
2. $\pi : \mathcal{S} \rightarrow \Pi \rightarrow \{0, 1\}$ is a valuation to propositional variables per state,
3. for all $1 \leq i \leq m$, $R_i \subseteq \mathcal{S} \times \mathcal{S}$ is an equivalence relation. For any $s \in \mathcal{S}$ and $i \leq m$, with $R_i(s)$ we mean $\{t \in \mathcal{S} \mid R_i st\}$.

We refer to the class of these Kripke models as $\mathcal{K}^m$, or, when m is understood, as $\mathcal{K}$.

Definition 2.5 The binary relation $\models$ between a formula φ and a pair $\mathcal{M}, s$ consisting of a model $\mathcal{M}$ and a state s in $\mathcal{M}$ is inductively defined by:

$$\begin{array}{ll}
\mathcal{M}, s \models p & \Leftrightarrow \pi(s)(p) = 1 \\
\mathcal{M}, s \models \varphi \wedge \psi & \Leftrightarrow \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}, s \models \psi \\
\mathcal{M}, s \models \neg\varphi & \Leftrightarrow \mathcal{M}, s \not\models \varphi \\
\mathcal{M}, s \models K_i\varphi & \Leftrightarrow \mathcal{M}, t \models \varphi \text{ for all } t \text{ with } (s,t) \in \mathrm{R}_i \\
\mathcal{M}, s \models G\varphi & \Leftrightarrow \mathcal{M}, t \models \varphi \text{ for all } t \text{ with } (s,t) \in \mathrm{R}_1 \cap \ldots \cap \mathrm{R}_m
\end{array}$$

For a given $\varphi \in \mathcal{L}$ and $\mathcal{M} \in \mathcal{K}$, $[\![\varphi]\!] = \{t \mid \mathcal{M}, t \models \varphi\}$. For sets of formulae Φ, $\mathcal{M}, s \models \Phi$ is defined by: $\mathcal{M}, s \models \Phi \Leftrightarrow \mathcal{M}, s \models \varphi$ for all $\varphi \in \Phi$. In this paper, a model $\mathcal{M}$ is always a $\mathcal{K}$-model.

The intuition behind the truth-definition of G is as follows: if t is a world which is not an epistemic alternative for agent i, then, if the agents would be able to communicate, all the agents would eliminate the state t. This is justified by the idea that the actual, or real state, is always an epistemic alternative for each agent (on $\mathcal{K}$-models, R_i is reflexive; or, speaking in terms of the corresponding axiom, knowledge is *veridical*). Using the wise-man metaphor: this man does not consider any state which has already been abandoned by one of the agents.

A formula φ is defined to be *valid* in a model $\mathcal{M}$ iff $\mathcal{M}, s \models \varphi$ for all $s \in \mathcal{S}$; φ is valid with respect to $\mathcal{K}$ iff $\mathcal{M} \models \varphi$ for all $\mathcal{M} \in \mathcal{K}$. The formula φ is *satisfiable* in $\mathcal{M}$ iff $\mathcal{M}, s \models \varphi$ for some $s \in \mathcal{S}$; φ is satisfiable with respect to $\mathcal{K}$ iff it is satisfiable in some $\mathcal{M} \in \mathcal{K}$. A set of formulae Φ is valid with respect to $\mathcal{M}/\mathcal{K}$ iff each formula $\varphi \in \Phi$ is valid with respect to $\mathcal{M}/\mathcal{K}$. We define two relations between a set of formulae and a formula: $\Phi \models^+ \varphi$ iff $\forall \mathcal{M}(\mathcal{M} \models \Phi \Rightarrow \mathcal{M} \models \varphi)$ and $\Phi \models^- \varphi$ iff $\forall \mathcal{M} \forall s(\mathcal{M}, s \models \Phi \Rightarrow \mathcal{M}, s \models \varphi)$.

Theorem 2.6 [Soundness and Completeness] Let φ be some formula, and let Φ be a set of formulae. The following soundness and completeness results hold.

- $\vdash \varphi \;\Leftrightarrow\; \models \varphi$
- $\Phi \vdash^+ \varphi \;\Leftrightarrow\; \Phi \models^+ \varphi$
- $\Phi \vdash^- \varphi \;\Leftrightarrow\; \Phi \models^- \varphi$

3 Group Knowledge is not always Implicit

As presented, within the system $\mathbf{S5}_m(G)$, all agents are considered 'equal': if we make no additional assumptions, they all know the same.

Lemma 3.1 For all $i, j \leq m$, we have: $\vdash K_i\varphi \Leftrightarrow \vdash K_j\varphi$.

Remark 3.2 One should be sensitive for the comment that Lemma 3.1 is a *meta statement* about the system $\mathbf{S5}_m(G)$, and as such it should be distinguished from the claim that one should be able to claim *within* the system $\mathbf{S5}_m(G)$ that all agents know the same: i.e., we do *not* have (nor wish to have) that $\vdash K_i\varphi \leftrightarrow K_j\varphi$, if $i \neq j$. Considering $\vdash \varphi$ as 'φ is derivable from an empty set of premises' (cf. our remarks following Definition 2.3), Lemma 3.1 expresses 'When no contingent fact is known on beforehand, all agents know the same'. In fact, one easily can prove that 'When no contingent facts are given, each agent knows exactly the $\mathbf{S5}_m(G)$- theorems', since we have, for each $i \leq m$: $\vdash \varphi \Leftrightarrow \vdash K_i\varphi$ ∎

Interestingly, we are able to prove a property like the one in Lemma 3.1 even when the G operator is involved.

Theorem 3.3 Let X and Y range over $\{K_1, K_2, \ldots, K_m, G\}$. Then: $\vdash X\varphi \Leftrightarrow \vdash Y\varphi$

Theorem 3.3 has, for both the reading as group knowledge as well as that of a receiving agent for G, some remarkable consequences. It implies that the knowledge in the group is nothing else than the knowledge of any particular agent. Phrased differently, let us agree upon what it means to say that group knowledge is implicit:

Definition 3.4 Given a notion of derivability $\vdash_a$, we say that group knowledge is *strongly implicit* (under $\vdash_a$) if there is a formula φ for which we have $\vdash_a G\varphi$, but for all $i \leq m \not\vdash_a K_i\varphi$. It is said to be *weakly implicit* under $\vdash_a$ if there is a set of premises Φ for which $\Phi \vdash_a G\varphi$, but $\Phi \not\vdash_a K_i\varphi$, for all $i \leq m$.

Thus, Theorem 3.3 tells us that, using $\vdash$ of $\mathbf{S5}_m(G)$, group knowledge is not implicit! Although counterintuitive at first, Theorem 3.3 also invites one to reconsider the meaning of $\vdash \varphi$ as opposed to a statement $\Phi \vdash^+ \varphi$. When interpreting the case where $\Phi = \emptyset$ as 'initially' or, more loosely, 'nothing has happened yet' (in particular when no communication of contingent facts has yet taken place), it is perhaps not too strange that all agents know the same and thus that all group knowledge is explicitly present in the knowledge of all agents.

With regard to derivability from premises, we have to distinguish the two kinds of derivation introduced in Definition 2.3.

Lemma 3.5 Let φ be a formula, Φ a set of formulae and i some agent.

- $\Phi \vdash^+ G\varphi \Leftrightarrow \Phi \vdash^+ K_i\varphi$
- $\Phi \vdash^- K_i\varphi \Rightarrow \Phi \vdash^- G\varphi$
- $\Phi \vdash^- G\varphi \not\Rightarrow \Phi \vdash^- K_i\varphi$

The first clause of Lemma 3.5 states that for derivability from premises in which necessitation on premises is allowed, group knowledge is also not implicitly but explicitly present in the individual knowledge of each agent in the group (thus Φ may be considered an initial set of facts that are known to everybody of the group). The second and third clause indicate that the notion of group knowledge is relevant only for derivation from premises in which necessitation on premises is not allowed. In that case, group knowledge has an additional value over individual knowledge. A further investigation into the nature of group knowledge when necessitation on premises is not allowed, is the subject of the next section.

4 Group Knowledge is not always Distributed

In the previous section we established that group knowledge is in some cases not implicit, but explicit. In particular for provability *per se* and for derivation from premises with necessitation on premises, group knowledge is rather uninteresting. The only case where group knowledge could be an interesting notion on its own, is that of derivability from premises without necessitation on premises. For this case, we want to formalize the notion of *distributed* knowledge. Although intuitively clear, a formalization of distributed knowledge brings a number of hidden parameters to surface. Informally, we say that the notion of group knowledge is *distributed* if the group knowledge (apprehended as a set of formulae) equals the set of formulae that can be derived from the union of the knowledge of the agents that together constitute the group.

The following quote is taken from a recent dissertation ([1]):

> *Implicit knowledge is of interest in connection with information dialogues: if we think of the dialog participants as agents with information states represented by epistemic formulae, then implicit knowledge precisely defines the propositions the participants could conclude to during an information dialogue ...*

Borghuis ([1]) means with 'implicit knowledge' what we call 'group knowledge'. We will see in this section, that using standard epistemic logic, one cannot guarantee that group knowledge is *precisely* that what can be concluded during an information dialogue.

To do so, we will first formalize the notion of distributivity: the following definition makes some of the hidden parameters explicit.

Definition 4.1 Let Φ denote a set of formulae, and φ a formula. Given two notions of derivability $\vdash_a$ and $\vdash_b$,

- we say that $\vdash_a$-derivable group knowledge is *strongly sound* with respect to $\vdash_b$-consequences, if, for all Φ and φ:

$$\Phi \vdash_a G\varphi \;\Rightarrow \exists\varphi_1,\ldots,\varphi_m: \; [\Phi \vdash_a K_1\varphi_1,\ldots,\Phi \vdash_a K_m\varphi_m \;\& \\ \Phi \vdash_b (\varphi_1 \wedge \ldots \wedge \varphi_m) \rightarrow \varphi] \quad (1)$$

- $\vdash_a$-derivable group knowledge is said to be *strongly complete* with respect to $\vdash_b$-consequences, if we have for all Φ and φ,

$$\Phi \vdash_a G\varphi \;\Leftarrow \exists\varphi_1,\ldots,\varphi_m: \; [\Phi \vdash_a K_1\varphi_1,\ldots,\Phi \vdash_a K_m\varphi_m \;\& \\ \Phi \vdash_b (\varphi_1 \wedge \ldots \wedge \varphi_m) \rightarrow \varphi] \quad (2)$$

- If both (1) and (2) hold for all Φ and φ, $\vdash_a$-derivable group knowledge is said to be *strongly distributed* (over the m agents) with respect to $\vdash_b$. If (1) only holds for $\Phi = \emptyset$, $\vdash_a$-derivable group knowledge is *weakly* sound with respect to $\vdash_b$-consequences; weak completeness and weak distributivity are defined similarly. If $\vdash_a$ or $\vdash_b$ is understood from context (for instance, if $\vdash_a=\vdash_b=$ $\mathbf{S5}_m(G)$), it is left out.

Note that $\vdash^+$ is both strongly and weakly distributed with respect to itself. Note furthermore that the definition of distributivity as given above already distinguishes sufficiently many parameters and is a fairly general one, of which we consider only specific cases in this paper. Still an even more general definition of distributivity could be given:

$$F(\Phi) \vdash_a G\varphi \;\Leftrightarrow \exists\varphi_1,\ldots,\varphi_m: \; [H_1(\Phi) \vdash_1 K_1\varphi_1,\ldots,H_m(\Phi) \vdash_m K_m\varphi_m \;\& \\ J(\Phi) \vdash_b (\varphi_1 \wedge \ldots \wedge \varphi_m) \rightarrow \varphi] \quad (3)$$

where $F, H_1, \ldots, H_m : \wp(\mathcal{L}) \to \wp(\mathcal{L})$ are functions that, for each agent, manipulate the set of premises Φ to a set that is allowed to be used by the agent. Moreover, J may restrict the premises to be used for the derivation of φ from the φ_i's. Thus, the scheme (3) allows for situations in which each agent has access to a specific part of the initial set of premises, and in which each of the agents uses its own logic. Moreover, it allows the derivation of φ from the formulae $\varphi_1, \ldots, \varphi_m$ to be done in a distinguished logic, possibly using some fixed set of formulae $J(\Phi)$.

Summarizing, scheme (3) is a general way to formalize our intuition that group knowledge is distributed if a formula is somehow declared to be group knowledge relative to some set of premises if and only if it follows in some sense (using some related set of premises) from the conjunction of m premises; and of each of these premises it can be somehow derived (again, using appropriate sets of formulae) that it is known by one of the agents.

Observation 4.2 Using only $\vdash^-$, group knowledge is strongly sound, but *not* strongly complete. However, under $\vdash^-$, group knowledge *is* weakly distributed.

Proof: The proof of the first observation is easy: from $\Phi \vdash^- G\varphi$ it follows by axiom $A8$ and Modus Ponens that $\Phi \vdash^- \varphi$. From this we derive $\Phi \vdash^- (\top \wedge \ldots \top) \to \varphi$. By m applications of $R2$ we have $\Phi \vdash^- K_1\top, \ldots, \Phi \vdash^- K_m\top$ which suffices to conclude that the claim indeed holds. Group knowledge is seen to be not complete under $\vdash^-$, by considering a suitable model, which we omit here. Finally, we argue that, under $\vdash^-$, group knowledge is weakly distributed. Weak soundness is implied by strong soundness, in order to prove weak completeness, since Φ is empty, we may use Theorem 3.3: thus $\vdash^- G\varphi$ implies $\vdash^- K_1\varphi$, which immediately yields the right hand side of (2). ∎

The proofs of the positive claims of Observation 4.2 tell us that they are obtained rather trivially. First of all, note that in order to prove weak distributivity, we used the fact that group knowledge is not implicit. (Thus, Lemma 3.5 also immediately gives weak soundness of group knowledge under $\vdash^+$ and $\vdash$.) Secondly, recall from the proof of Observation 4.2 that, using only $\vdash^-$, group knowledge is trivially proven to be strongly sound: If $G\varphi$ follows from a given set of true premises Φ, the formula φ itself follows from this set Φ, and there is no point in seeking for 'reasons' for this formula φ in the knowledge of the individual agents. It seems more reasonable to require that φ follows *independently from the given set of premises* from a set of formulae known by the agents:

$$\Phi \vdash^- G\varphi \Rightarrow \exists \varphi_1, \ldots, \varphi_m : [\Phi \vdash^- K_1\varphi_1, \ldots, \Phi \vdash^- K_m\varphi_m \;\&\; \vdash^- (\varphi_1 \wedge \ldots \wedge \varphi_m) \to \varphi] \quad (4)$$

Recall from Definition 2.3 that in fact there is no point, nor any harm, in writing the $^-$ with the last occurrence of '$\vdash$', if there are no premises involved, we may just write '$\vdash$'. Thus (4) expresses the fact that, under $\vdash^-$, group knowledge is strongly sound under plain $\mathbf{S5}_m(G)$-consequences. Let us try and rephrase equation (4) semantically.

Definition 4.3 Let $\mathcal{M}$ be a Kripke model with state s, and $i \leq m$. The *knowledge set* of i in $\mathcal{M}, s$ is defined by: $K(i, \mathcal{M}, s) = \{\varphi \mid \mathcal{M}, s \models K_i\varphi\}$.

Now, the semantic counterpart of (4) reads as follows:

$$\mathcal{M}, s \models G\varphi \;\Rightarrow\; \bigcup_i K(i, \mathcal{M}, s) \models^- \varphi \quad (5)$$

Property (5) expresses that G-knowledge can only be true at some world s if it is derivable from the set that results when putting all the knowledge of all the agents in s together.

For multi-agent architectures in which agents have the possibility to communicate (like for instance [8]) the principle (5) is rather relevant. This so called *principle of full communication* captures the intuition of *fact discovery* (cf. [4]) through communication. The principle of full communication formalizes the intuitive idea that it is possible for one agent to become a wise man by communicating with other agents: these other agents may pass on formulae from their knowledge that the receiving agent combines to end up with the knowledge that previously was implicit. As such, the principle of full communication seems highly desirable a property for group knowledge. It is questionable whether group knowledge is of any use if it cannot somehow be upgraded to explicit knowledge by a suitable combination of the agents' individual knowledge sets, probably brought together through communication. Coming back to the example of the three computer scientists and the question whether $P \neq NP$; if there is no way for them to combine their knowledge such that the proof results, it is not clear whether this proof should be said to be distributed over the group at all.

Unfortunately, in the context of $\mathbf{S5}_m(G)$, the principle of full communication does not hold. The following counterexample describes a situation in which group knowledge cannot be upgraded to explicit knowledge (thereby answering a question raised in [8]).

Counterexample 4.4 Let the set Π of propositional variables be given by the singleton set $\{q\}$. Consider the Kripke model $\mathcal{M}$ such that

- $\mathcal{S} = \{x_1, x_2, y_1, y_2, z_1, z_2\}$,
- $\pi(q, x_j) = \mathbf{1}, \pi(q, y_j) = \mathbf{1}, \pi(q, z_j) = \mathbf{0}$ for $j = 1, 2$,
- R_1 is given by the solid lines in figure 1, and R_2 is given by the dashed ones.

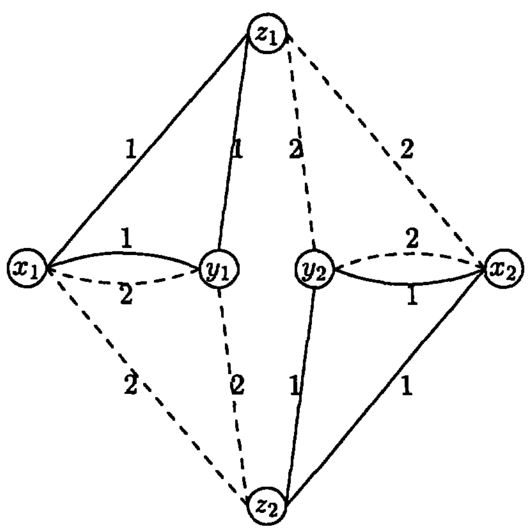

Figure 1: The epistemic accessibility relations.

Observation 4.5 It holds that:

- $K(1, \mathcal{M}, x_1) = K(2, \mathcal{M}, x_1)$
- $\mathcal{M}, x_1 \models Gq$
- $q \notin K(1, \mathcal{M}, x_1) \cup K(2, \mathcal{M}, x_1) = K(1, \mathcal{M}, x_1)$

From this last observation it follows that the principle of full communication does not hold in this model: although the formula q is group knowledge, it is not possible to derive this formula from the combined knowledge of the agents 1 and 2. Thus it is not possible for one of these agents to become a wise man through receiving knowledge from the other agent. In terms of Borghuis: the model of Counterexample 4.4 shows that one can have group knowledge of an atomic fact q, although this group knowledge will never be derived during a dialogue between the agents that are involved.

Corollary 4.6 Using $\vdash^-$, group knowledge is not strongly sound with respect to $\mathbf{S5}_m(G)$-consequences, i.e., property (5) does not hold.

Summarizing, we have seen that $\vdash^-$-derivable group knowledge is weakly distributive over $\vdash^-$-consequences in a trivial manner, but it is not sound with respect to a more interesting notion of derivability: it is not always the case that group knowledge is derivable from all the agent's knowledge together.

5 Models in which Group Knowledge is Distributed

In this section, we will characterize a class of Kripke models, in which G-knowledge is always distributed. To be more precise, the models that we come up with will satisfy

$$\mathcal{M}, s \models G\varphi \Leftrightarrow \bigcup_i K(i, \mathcal{M}, s) \models^- \varphi \quad (6)$$

Definition 5.1 A Kripke model $\mathcal{M} = \langle \mathcal{S}, \pi, \mathrm{R}_1, \ldots, \mathrm{R}_m \rangle$ is called *finite* if $\mathcal{S}$ is finite; moreover it is called a *distinguishing model* if for all $s, t \in \mathcal{S}$ with $s \neq t$, there is a φ_{s^+,t^-} such that $\mathcal{M}, s \models \varphi_{s^+,t^-}$ and $\mathcal{M}, t \not\models \varphi_{s^+,t^-}$.

When considering an $\mathbf{S5}_1$–model as an epistemic state of an agent, it is quite natural to use only distinguishing models: such a model comprises all the different possibilities the agent has. One can show that in the 1-agent case, questions about satisfiability of finite sets of formulae, and hence that of logical consequence of a finite set of premises, can be decided by considering only finite distinguishing models. In fact, one only needs to require such models to be distinguishing at the propositional level already: any two states in such a *simple* **S5**-model differ in assigning a truth value to at least one atom (see, e.g. [7], Section 1.7). In the multiple-agent case, this distinguishing requirement needs to be lifted from the propositional level to the whole language $\mathcal{L}$.

The nice feature of finite distinguishing models is that sets of states can be named:

Lemma 5.2 Let $\mathcal{M} = \langle \mathcal{S}, \ldots \rangle$ be a finite distinguishing model. Then:

$$\forall X \subset \mathcal{S}\ \exists \alpha_X \in \mathcal{L}\ \forall x \in \mathcal{S}\ \ (\mathcal{M}, x \models \alpha_X \Leftrightarrow x \in X)$$

For a given set X, we call α_X the *characteristic formula for* X.

Lemma 5.3 Let $\mathcal{M} = \langle \mathcal{S}, \pi, R_1, \ldots, R_m \rangle$ be given, with $s \in \mathcal{S}$. Suppose that $Z \subseteq \mathcal{S}$ is such that for all $z \in Z$ we have $\mathcal{M}, z \models \zeta$. Moreover, suppose $X_1, \ldots, X_n \subseteq \mathcal{S}$ are such that $R_i(s) \subseteq (X_1 \cup \ldots \cup X_n \cup Z)$. Then

$$\mathcal{M}, s \models K_i(\neg\alpha_{X_1} \to (\neg\alpha_{X_2} \to (\ldots(\neg\alpha_{X_n} \to \zeta)\ldots)))$$