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

Wiebe van der Hoek
Bernd van Linder
John-Jules 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 this definition gives rise to some quite counter-intuitive behaviour. 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 fulfil 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. [3, 14]). 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 G 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'

*This is an improved and extended version of [6] 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 $\psi_3 = (\psi_1 \wedge \psi_2) \rightarrow \varphi$. 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 G 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 [3]. 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 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 G 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. Proofs of theorems are provided in the appendix.

2 Knowledge and Group Knowledge

Halpern and Moses introduced an operator to model group knowledge ([3]). 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. [10]), later on, the term 'distributed knowledge' (with the operator D) became the preferred name for the group knowledge we want to consider here (cf. [4]). 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 [3]; a system to model such communication was proposed in [11]). 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 $\mathbf{S5}$ logic incorporating the operator G.

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

$A1$	any axiomatization for propositional logic
$A2$	$(K_i\varphi \wedge K_i(\varphi \to \psi)) \to K_i\psi$
$A3$	$K_i\varphi \to \varphi$
$A4$	$K_i\varphi \to K_iK_i\varphi$
$A5$	$\neg K_i\varphi \to K_i\neg K_i\varphi$
$A6$	$K_i\varphi \to G\varphi$
$A7$	$(G\varphi \wedge G(\varphi \to \psi)) \to G\psi$
$A8$	$G\varphi \to \varphi$
$A9$	$G\varphi \to GG\varphi$
$A10$	$\neg G\varphi \to G\neg G\varphi$

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

$R1$	$\vdash \varphi, \vdash \varphi \to \psi \;\Rightarrow \vdash \psi$
$R2$	$\vdash \varphi \;\Rightarrow \vdash K_i\varphi$, 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 $R2$.

To relate this notion of group knowledge to the motivating example of Section 1, in which the group is able to derive a conclusion φ from the lemma's ψ_1, ψ_2 and ψ_3, each known by one of the agents, let us consider the following derivation rule which is to be understood to hold for all $\varphi, \psi_i, i \leq m$

$$R3 \quad \vdash (\psi_1 \wedge \cdots \wedge \psi_m) \to \varphi \Rightarrow \vdash (K_1\psi_1 \wedge \cdots \wedge K_m\psi_m) \to G\varphi$$

Theorem 2.3 Let the logic $\mathbf{S5}_m(G)[R3/A6]$ be the as $\mathbf{S5}_m(G)$, with axiom $A6$ replaced by the rule $R3$. Then, the two systems are equally strong: for all φ, $\mathbf{S5}_m(G) \vdash \varphi \Leftrightarrow \mathbf{S5}_m(G)[R3/A6] \vdash \varphi$.

From the proof of this theorem, as given in [14, page 263] one even can derive that this equivalence is obtained for any multi-modal logic that is normal with respect to G, i.e., for any logic $\mathbf{X}$ containing the axioms $A7$ and the rules $R1$ and $R2' : \vdash \varphi \Rightarrow \vdash G\varphi$ we have an equivalence between $\mathbf{X} + A6$ and $\mathbf{X} + R3$.

We now define two variants of *provability from premises*: one in which necessitation on premises is allowed and one in which it is not.

Definition 2.4 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 $\varphi_j = \varphi_k \to \varphi_i$
- 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:

$$\emptyset \vdash^+ \varphi \Leftrightarrow \emptyset \vdash^- \varphi$$

such that $\vdash \varphi$ is unambiguously defined as $\emptyset \vdash^+ \varphi$ or $\emptyset \vdash^- \varphi$, and

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

Also, one can prove the following connection between the two notions of derivability from premises.

Theorem 2.5 *Let $\{K_1, \ldots, K_m\}^*$ be the set of sequences of epistemic operators. Then*
$\{X\varphi \mid X \in \{K_1, \ldots, K_m\}^*, \varphi \in \Phi\} \vdash^- \psi \;\Leftrightarrow\; \Phi \vdash^+ \psi$

From [13, Lemma A1], for the one agent case without group knowledge (i.e., $\mathbf{S5}_1$), one even obtains

$$\{K\varphi \mid \varphi \in \Phi\} \vdash^- \psi \Leftrightarrow \Phi \vdash^+ \psi$$

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 2.6 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. $\pi : \mathcal{S} \to \Pi \to \{\mathbf{0}, \mathbf{1}\}$ 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 $\mathbf{S5}^m$, or, when m is understood, as $\mathbf{S5}$.

Definition 2.7 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:

$$\begin{array}{lll}
\mathcal{M}, s \models p & \Leftrightarrow & \pi(s)(p) = \mathbf{1} \\
\mathcal{M}, s \models \varphi \wedge \psi & \Leftrightarrow & \mathcal{M}, s \models \varphi \text{ and } \mathcal{M}, s \models \psi \\
\mathcal{M}, s \models \neg\varphi & \Leftrightarrow & \mathcal{M}, s \not\models \varphi \\
\mathcal{M}, s \models K_i\varphi & \Leftrightarrow & \mathcal{M}, t \models \varphi \text{ for all } t \text{ with } (s,t) \in R_i \\
\mathcal{M}, s \models G\varphi & \Leftrightarrow & \mathcal{M}, t \models \varphi \text{ for all } t \text{ with } (s,t) \in R_1 \cap \ldots \cap R_m
\end{array}$$

For a given $\varphi \in \mathcal{L}$ and $\mathcal{M} \in \mathbf{S5}$, $[\![\varphi]\!] = \{t \mid \mathcal{M}, t \models \varphi\}$. 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 $\mathbf{S5}$-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 **S5**-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 **S5** iff $\mathcal{M} \models \varphi$ for all $\mathcal{M} \in$ **S5**. 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 **S5** iff it is satisfiable in some $\mathcal{M} \in$ **S5**. A set of formulae Φ is valid with respect to $\mathcal{M}$/**S5** iff each formula $\varphi \in \Phi$ is valid with respect to $\mathcal{M}$/**S5**. 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 2.8 [Soundness and Completeness] Let φ be some formula, and let Φ be a set of formulae. The following soundness and completeness results hold.

1. $\vdash \varphi \;\Leftrightarrow \models \varphi$
2. $\Phi \vdash^{-} \varphi \;\Leftrightarrow\; \Phi \models^{-} \varphi$
3. $\Phi \vdash^{+} \varphi \;\Leftrightarrow\; \Phi \models^{+} \varphi$

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 2.4), 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 2.4.

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 using the notions of derivability used so far. The reader may wish to skip Section 4.1 and move on to the semantic approach in Section 4.2 at a first reading.

4.1 A deductive approach

To formalise what we mean with distributed group knowledge, let us have a second look at derivation rule $R3$ first. It says that, if φ follows from the conjunction of $\varphi_1, \ldots, \varphi_m$, then, if we also have $K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m$, we are allowed to conclude that φ is known by the group, i.e. then also $G\varphi$. Now, we are inclined to call group knowledge *distributed* if the converse also holds, that is, if everything that is known by the group does indeed follow from the knowledge of the group's individuals. Let us, for convenience, say that group knowledge is *perfect* (over the individuals) if both directions hold. What we now have to do is become precise about the phrase 'follows from' and about the role that can be played be premises here. Let us summarize our first approximation to the crucial notions in a semi formal definition:

Definition 4.1 (First approximation).

1. We say that G represents *complete* group knowledge if we have that for every φ and $\varphi_1, \ldots, \varphi_m$ such that $(K_1\varphi \wedge \cdots \wedge K_m\varphi_m)$ can be derived, and also the implication $(\varphi_1 \wedge \cdots \wedge \varphi_m) \to \varphi$, then we also have that $G\varphi$ can be derived.

2. The group knowledge represented by G is called *distributed* if for every φ such that $G\varphi$ is derived we can find $\varphi_1, \ldots, \varphi_m$ such that: (1) together they imply φ and (2) The conjunction $(K_1\varphi \wedge \cdots \wedge K_m\varphi)$ is derivable.

3. If both items above hold we call the group knowledge represented by G *perfect group knowledge*.

Observation 4.2 In $\mathbf{S5}_m(G)$, using $\vdash$, G-knowledge is both complete group knowledge and distributed.

The proof of Observation 4.2 shows that it holds for a trivial reason. For distributivity, if $\vdash G\varphi$, by Theorem 3.3, we also have $\vdash K_1\varphi, \ldots, \vdash K_m\varphi$, and then distributivity follows from $\vdash (\varphi \wedge \cdots \wedge \varphi) \rightarrow \varphi$. To show that G models complete group knowledge, one reasons as follows: from $\vdash K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m$ and $\vdash (\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow \varphi$ one obtains $\vdash G\varphi_1 \wedge \cdots \wedge G\varphi_m$ and $\vdash G((\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow \varphi)$ from which then $\vdash G\varphi$ immediately follows.

So, to obtain distributivity in a non-trivial way, we should allow for premises. However, we know from the previous section that we then have at least two notions of derivability to consider. Unfortunately, as the following observation summarizes, there is no simple choice here.

Observation 4.3 If the notion of derivability used in Definition 4.1 is to be $\Phi \vdash^{+} \psi$ or $\Phi \vdash^{-} \psi$, we again obtain distributivity of G-knowledge in a trivial way.

Let us try to understand the source of Observation 4.3. For both $\star = +, \Leftrightarrow$, if $\Phi \vdash^{\star} G\varphi$, then we also have $\Phi \vdash^{\star} \varphi$, and hence $\Phi \vdash^{\star} (\top \wedge \cdots \top) \rightarrow \varphi$. Obviously, we also have $\Phi \vdash^{\star} (K_1\top \wedge \cdots \wedge K_m\top)$, justifying the claim that group knowledge is distributed.

From the two Observations 4.2 and 4.3 above, we conclude that, in order to have an interesting notion of distributed group knowledge, we have to allow for a subtle use of premises and of notions of derivability. Let us therefore introduce the following notation. Let $\kappa \in \{0, 1\}$ be a characteristic function in the sense that, for every set Φ, $\kappa \cdot \Phi = \emptyset$ if $\kappa = 0$, and Φ else. Now, we want to distinguish three types of derivability: one to obtain the $(K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$ formulas (the 'knowledge conclusions'), one to obtain $(\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow \varphi$ (the 'logical conclusions') and one to obtain formulas of type $G\varphi$ (the 'group knowledge conclusions'). For each type of derivability, we may choose to allow for premises or not, and we also choose between $\vdash^{+}$ and $\vdash^{-}$.

Definition 4.1 Let $\kappa_k, \kappa_l, \kappa_g \in \{0, 1\}$ and let K, L and G be variable over $\{+, \Leftrightarrow\}$. For each tuple $\Lambda = \langle \kappa_k, K, \kappa_l, L, \kappa_g, G\rangle$ we now define the following notions:

- We say that Λ-derivable group knowledge is complete if, for all sets of premises Φ and every formula φ:

$$\kappa_g \cdot \Phi \vdash^{G} G\varphi \quad \Leftarrow \exists \varphi_1, \ldots, \varphi_m : \quad [\kappa_k \cdot \Phi \vdash^{K} (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m) \& \\ \kappa_l \cdot \Phi \vdash^{L} (\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow \varphi] \quad (1)$$

- we say that Λ-derivable group knowledge is distributed, if for every set of premises Φ, and every φ:

$$\kappa_g \cdot \Phi \vdash^G G\varphi \quad \Rightarrow \exists \varphi_1, \ldots, \varphi_m : \quad [\kappa_k \cdot \Phi \vdash^K (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m) \\ \kappa_l \cdot \Phi \vdash^L (\varphi_1 \wedge \ldots \wedge \varphi_m) \rightarrow \varphi] \quad (2)$$

- If both (1) and (2) hold for a tuple Λ and every Φ and φ, we say that the knowledge represented by G is perfect with respect to Λ.

Thus, we have specialized our first attempt to define perfect group knowledge, in notions depending on the choices in Λ. There are in principle 64 of such notions, but the following observations give a systematic account of them: the reader may also consult the summarising table Table 1 in the Appendix.

Observation 4.4 For 49 possibilities for $\Lambda = \langle \kappa_k, K, \kappa_l, L, \kappa_g, G \rangle$ one can show that group knowledge is indeed distributed—but the proofs show that this is only established in a 'trivial' manner.

In the observation above, 'trivial' refers to the fact distributivity is proven in a way similar to that in Observation 4.2 and 4.3–for an exhaustive proof, the reader is referred to the appendix. Those cases are summarized in Table 1, with an a, b c or d in the 'yes' column.

Observation 4.5 For the 13 of the remaining 15 possibilities for $\Lambda = \langle \kappa_k, K, \kappa_l, L, \kappa_g, G \rangle$ one can show that group knowledge is not distributed; they are summarized in Table 1 with the items e, f and g in the 'no' column.

The remaining interesting cases are $\Lambda = \langle 1, \Leftrightarrow, 0, L, \kappa_1, 1 \rangle$ which we take as a starting point for the next section. Note that in fact this is only one case, since if $\kappa_l = 0$, the choice of $L = \Leftrightarrow, +$ is arbitrary. Also note that, apart from the technical results of the observations above, this remaining Λ seems to make sense: we do not distinguish between κ_k and $\kappa_g (= 1)$, we take premises seriously but we only consider the case where $K = G = \Leftrightarrow$, so that group knowledge can be called implicit, according to Section 3. Then, according to Observation 4.3 the only sensible choice for κ_l is 0. Let us formulate the notion of distributivity for the remaining Λ's:

$$\Phi \vdash^- G\varphi \quad \Rightarrow \exists \varphi_1, \ldots, \varphi_m : \quad \vdash^K (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m) \\ \Phi \vdash^- (\varphi_1 \wedge \ldots \wedge \varphi_m) \rightarrow \varphi] \qquad (3)$$

Thus, the conclusion φ should be an $\mathbf{S5}_m(G)$-consequence of the lemma's $\varphi_1, \ldots, \varphi_m$, for $G\varphi$ and $(K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$ that are derivable using the same rules: using premises but without applying necessitation to them.

4.2 A semantic approach

If we try to reformulate (3) semantically, we obtain

> *If $G\varphi$ is true in every situation that verifies Φ, then there should be $\varphi_1, \ldots, \varphi_m$ such that $(\varphi_1 \wedge \cdots \wedge \varphi_m) \to \varphi$ is valid and $(K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$ is true in every situation verifying Φ.*

However, we think that the order of quantification in the citation above deserves some more attention. Consider $\Phi = \{K_1p \vee K_2p\}$. Indeed, we have $\Phi \models^- Gp$, but does p follow from formulas φ_1, φ_2 such that $(K_1\varphi \wedge K_2\varphi_2)$ is true in every situation $\mathcal{M}, s$ for Φ? No, obviously not (see Observation 4.6 below or its proof in the Appendix); this is too strong a requirement: instead of finding such φ_1 and φ_2 such that every situation $\mathcal{M}, s$ for Φ satisfies $(K_1\varphi \wedge K_2\varphi_2)$, however, we are able to find such formula for every situation that verifies Φ, since such a situation $\mathcal{M}, s$ either satisfies K_1p or K_2p, so that we can take (φ_1, φ_2) to be either $(p, \top)$ or $(\top, p)$.

Observation 4.6 Although we have $\{K_1p \vee K_2p\} \vdash^- Gp$, there exist no φ_1, φ_2 with $\vdash (\varphi_1 \wedge \varphi_2) \to p$ and $\{K_1p \vee K_2p\} \vdash^- (K_1\varphi_2 \wedge K_2\varphi_2)$.

Corollary 4.7 Group knowledge is not distributed in the sense of (3).

So, according to the latter corollary, also the remaining choices for Λ do not yield a notion of distributed knowledge. Let us thus put our argument about the order of quantification one step further:

> *We say that group knowledge satisfies the principle of full communication if in every situation in which $G\varphi$ is true one can find $\varphi_1, \ldots, \varphi_m$ such that $(K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$ also holds in that situation, and, moreover, $(\varphi_1 \wedge \cdots \wedge \varphi_m) \to \varphi$ is a valid implication.*

Phrased negatively, group knowledge does not satisfy the principle of full communication if there is a situation in which the group knows something i.e., $G\varphi$, where φ does not follow from the knowledge of the individual group members.

Definition 4.8 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 full communication reads as follows:

$$\mathcal{M}, s \models G\varphi \;\Rightarrow\; \bigcup_i K(i, \mathcal{M}, s) \models^- \varphi \tag{4}$$

Property (4) 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 [11]) the principle (4) is rather relevant. This so called *principle of full communication* captures the intuition of *fact discovery* (cf. [5]) 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 statement 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 [11]).

Counterexample 4.9 Let the set Π of propositional variables contain the atom 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 $j = 1, 2$ and for all $p \in \Pi$, $\pi(p, x_1) = \pi(p, x_2)$, $\pi(p, y_1) = \pi(p, y_2)$ and $\pi(p, z_1) = \pi(p, z_2)$,

- R_1 is given by the solid lines in Figure 1, and R_2 is given by the dashed ones (reflexive arrows are omitted).

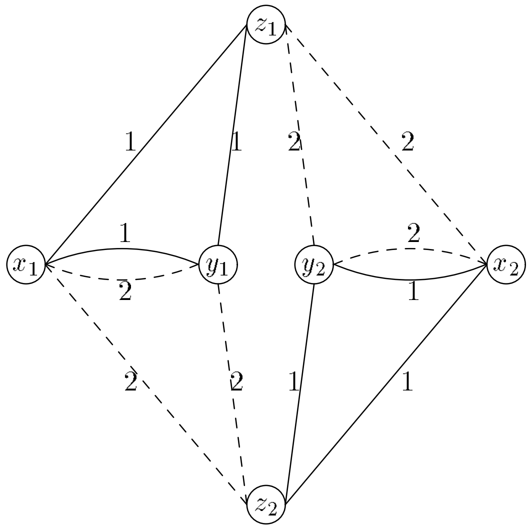

Figure 1: The epistemic accessibility relations.

Observation 4.10 It holds that (see Appendix A for a justification):

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

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 4.9[1] 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.

[1] Mark Ryan correctly pointed out to us that in fact there is a smaller model that proves our case: the y-worlds are in fact superfluous. We stick to our original model in order not to suggest that one has to adopt $G\varphi \leftrightarrow \varphi$ to be able to make Observation 4.10.

Corollary 4.11 In $\mathbf{S5}_m(G)$, group knowledge does not satisfy the principle of full communication

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, in the sense that they satisfy the principle of full communication. To be more precise, the models that we come up with will satisfy

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

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. [14], 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:

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

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 a given finite distinguishing model, 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

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

If one tries to link up the two lemma's above with the principle of full communication, one may take the following view point. First of all, note that, given a state s, all agents at least share one epistemic alternative, which is s itself. Now, consider the set of epistemic alternatives for agent i in s, $R_i(s)$. Relative to agent j we can partition $R_i(s)$ in two subsets: a set $Y_j \subseteq R_i(s)$ of worlds that are considered as epistemic alternatives for agent j, and a set $X_j = R_i(s) \setminus Y_j$ of worlds that are possible alternatives for i, but not for j. Let α_{X_j} and α_{Y_j} be the characteristic formulas for X_j and Y_j. respectively. Then, agent i knows that the real world must be in either one of X_j and Y_j, i.e. agent i knows that $\neg\alpha_{X_j} \to \alpha_{Y_j}$, and agent j knows $\neg\alpha_{X_j}$. Thus, agent j may inform agent i so to speak which of i's alternatives can be given up. Now, if all the agents h would communicate this information X_h to agent i, then finally agent i would end up with those epistemic alternatives that are considered possible by all of the agents. This argument is made more precise in the (proof of) the following theorem.

Theorem 5.4 Let $\mathcal{M}$ be a finite distinguishing model. Then, in $\mathcal{M}$, G-knowledge is distributed, in the sense that it satisfies (5).

We finally observe that the two properties of Definition 5.1 are in some sense also necessary for distributivity. Let us say that in $\mathcal{M}$ group knowledge satisfies the principle of full communication if, for all states s and formulae φ, equation (5) is satisfied. Then:

Observation 5.5

- There is a finite, not distinguishing model, in which group knowledge does not fulfil the principle of full communication.
- There is a distinguishing model, not finite, in which group knowledge does not fulfil the principle of full communication.

6 Conclusion

We have given formal definitions of what it means for group knowledge to be implicit or to be distributed relative to a group. Basically, group knowledge is implicit, if there is some formula known by the group, but not by any of its members. Here, one has to be precise about the role to be played by premises, and about the notion of derivability that is used to conclude that something is known. It appears that, under the standard notion of derivability in $\mathbf{S5}_m(G)$, and when no premises are allowed, group knowledge is not implicit. The same holds if one allows for applications of the Necessitation rule to premises. Thus, the only case in which group knowledge may be implicit in a group, is the case in which premises describe a situation, rather than the knowledge present beforehand.

Then, we studied whether group knowledge was always distributed. We made the notion of distributivity precise by saying that a formula is distributed (given some set of premises) if it can somehow, using some notion of derivability and some related set of premises, be derived from the union of the agents' knowledge. We argued that there was one way of choosing the parameters such that distributivity is not trivialised. We then gave a semantic account of a weak variant of distributivity (called the principle of full communication): we then were able to show that principle was still not guaranteed. Then, we gave two properties on Kripke models that guarantee that group knowledge *does* allow full communication; we also showed that both properties are, in some sense, needed.

We think it quite remarkable that such an appealing and natural notion like group knowledge as discussed in this paper, can give rise to rather complicated behaviour or properties. For instance, it is also well known that the canonical models for a logic like $\mathbf{S5}_m(G)$ only have the property that the accessibility relation for G, say R_G is a *subset* of $R_1 \cap \ldots \cap R_m$, and cannot be guaranteed to be equal to this intersection. In other words, this equality is not modally definable, although one can find a complete axiomatization for those models that satisfy it. For more on this, the reader is referred to [7]. In that paper, the authors also show that definability of $R_1 \cap \ldots \cap R_m = R_G$ *is* achieved when using graded modal logic. But in this logic, the assumption of distinguishability of a model seems to be unnatural; graded modal logic is typically designed to be able to reason about numbers (and hence copies) of worlds.

Also, in [8] the same authors demonstrate that proving an epistemic logic involving both group knowledge and common knowledge complete with respect to $\mathbf{S5}_m(G)$-like models can be quite complicated. Again, distinguishability of the models have to be given up. To be more precise, given a state $\mathcal{M}, s$ for ψ, the technique of *filtration* in modal semantics often yields a finite distinguishing model $\mathcal{M}'$, such that $\mathcal{M}', s'$ still verifies ψ, which might suggest af first sight that finite distinguishing models are always obtained. It is not guaranteed, however, that this model $\mathcal{M}$ is in the appropriate class, i.e., if $\mathcal{M}$ is an $\mathbf{S5}_m(G)$-models, $\mathcal{M}'$ need not be.

Finally, in a recent manuscript ([12]) a rather surprising property of group knowledge was mentioned in the context of so-called *Interpreted Systems*, systems that are widely accepted as a general semantics for distributed systems. In such a system, one obtains $G\varphi \leftrightarrow \varphi$, for all φ. Since such a system is by definition distinguishing (the states usually differ already on the objective level), in the finite versions the principle of full communication would be automatically obtained, but, at the same time, it might be a too strong, and perhaps optimistic principle: all facts about the world would be eventually obtained by the agents, would they be able to fully communicate.

Acknowledgements This work was partially supported by ESPRIT WG No. 23531 (FIREworks) and the Human Capital and Mobility network EXPRESS. Mark Ryan and Elias Thijsse gave useful comments on previous versions of this paper. We also thank the participants of LOFT2 (the second conference on Logic and the Foundations of the Theory of games and decisions) for the inspiring discussion on this paper.

References

[1] T. Borghuis. *Coming to Terms with Modal Logic: On the interpretation of modalities in typed λ-calculus.* PhD thesis, Technical University Eindhoven, 1994.

[2] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge.* MIT Press, Cambridge Massachusetts, 1995.

[3] J.Y. Halpern and Y. Moses. A guide to the modal logics of knowledge and belief. In *Proc. IJCAI-85*, pages 480–490, Los Angeles, CA, 1985.

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

[5] J.Y. Halpern and Y.O. Moses. Knowledge and common knowledge in a distributed environment. *Journal of the ACM*, 37(3):549–587, 1990.

[6] W. van der Hoek, B. van Linder, and J.-J.Ch. Meyer. Group knowledge isn't always distributed. In *Proceedings of the Bi-annual Israelian Symposium on the Foundations of AI '95*, pages 191–200, 1995.

[7] W. van der Hoek and J.-J. Ch Meyer. Making some issues of implicit knowledge explicit. *International Journal on Foundations of Computer Science*, 3(2):193–224, 1992.

[8] W. van der Hoek and J.-J.Ch. Meyer. A complete epistemic logic for multiple agents: Combining distributed and common knowledge. Technical Report UU-CS-1996-52, Utrecht University, 1995.

[9] G.E. Hughes and M.J. Cresswell. *A Companion to Modal Logic*. Methuen, 1984.

[10] H. Levesque. A logic of implicit and explicit belief. In *Proceedings of the National Conference on Artificial Intelligence*, pages 198–202, 1984.

[11] B. van Linder, W. van der Hoek, and J.-J.Ch. Meyer. Communicating rational agents. In B. Nebel and L. Dreschler-Fischer, editors, *KI-94: Advances in Artificial Intelligence*, volume 861 of *Lecture Notes in Computer Science (subseries LNAI)*, pages 202–213. Springer-Verlag, 1994.

[12] A. Lomuscio and M. Ryan. A note on the relation between interpreted systems and kripke models. unpublished manuscript, 1997.

[13] J.-J. Ch. Meyer and W. van der Hoek. A default logic based on epistemic states. *Fundamentae Informatica*, 23(1):33–65, 1995.

[14] J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Number 41 in Cambridge Tracks in Theoretical Computer Science. Cambridge University Press, 1995.

A Proofs

Proof of Theorem2.5 Our proof is inspired by that of [13, LEMMA A1], where a similar property is proven for an '$\mathbf{S5}_1$-like' logic. Let us abbreviate $\{X \mid X \in \{K_1, \ldots K_m\}^+$ to W (the set of non-empty words over the alphabet $\{K_1, \ldots K_m\}$) and, for a set of premises Φ, let us write $W\Phi$ for $\{X\varphi \mid X \in W, \varphi \in \Phi\}$. We first prove the following lemma

Lemma A.1 *For all* $\Phi, \varphi, i \leq m$*:* $W\Phi \vdash^- \psi \Leftrightarrow W\Phi \vdash^- K_i\psi$

Proof: The '$\Leftarrow$' direction follows immediately from an application of $R1$ to $K_i\psi$ and axiom $A3$. The '$\Rightarrow$' direction is proven by induction on the length $\ell(\pi^-)$ of a shortest $\vdash^-$-proof $\pi = \langle\psi_1, \ldots \psi_{\ell(\pi)} = \psi\rangle$ of ψ from $W\Phi$, i.e. we show that, for all n and all ψ,

$$\text{If } W\Phi \vdash^- \psi \text{ has a shortest proof of length } n, \text{ then } W\Phi \vdash^- K_i\psi \quad (6)$$

- If $\ell(\pi) = 1$ then
  1. either $\psi \in W\Phi$, i.e., $\psi = X\varphi$ for some sequence $X = K_{j_1}K_{j_2} \ldots K_{j_t} \in W$ and $\varphi \in \Phi$. Obviously, $K_iX\varphi \in W\Phi$, and hence $W\Phi \vdash^- K_i\psi$.
  2. Or ψ is an axiom of $\mathbf{S5}_m(G)$. By necessitation, we also have $\vdash K_i\psi$ and thus $W\Phi \vdash^- K_i\psi$.
- Suppose that (6) holds for all $W\Phi \vdash^- \alpha$ with $\ell(\pi) \leq k (k \geq 1)$ and suppose now that $W\Phi \vdash^- \psi$ has a shortest proof $\pi = \langle\psi_1, \ldots \psi_{k+1}(= \psi)\rangle$.
  1. We don't have to consider the case that $\psi \in W\Phi$, since then π would not be a shortest proof. The same holds for the case ψ being an $\mathbf{S5}_m(G)$-axiom.
  2. Suppose $\psi = K_j\beta$ with $\vdash \beta$. By necessitation, we then also have $\vdash K_i\beta$ and we prove $K_i\psi = K_iK_i\beta$ by applying $A4$ and $R1$, yielding $\vdash K_i\psi$ and thus $W\Phi \vdash^- K_i\psi$.
  3. The only other possibility is that ψ_k is obtained by an application of modus ponens to ψ_j and $\psi_j \rightarrow \psi_{k+1}$, $j \leq k$. Using the induction hypothesis on the proofs for ψ_j and $\psi_j \rightarrow \psi_{k+1}$, we obtain $W\Phi \vdash^- K_i\psi_j$, $W\Phi \vdash^- K_i(\psi \rightarrow \psi_{k+1})$. Applying $R1$ and $A2$ to the latter two conclusions yields $W\Phi \vdash^- K_i\psi_{k+1}$, i.e. $W\Phi \vdash^- K_i\psi$.

■

We now reformulate and proof Theorem 2.5:

Theorem 2.5 Let $\{K_1, \dots K_m\}^+$ be the set of non-empty sequences of epistemic operators. Then
$\{X\varphi \mid X \in \{K_1, \dots K_m\}^+, \varphi \in \Phi\} \vdash^- \psi \;\Leftrightarrow\; \Phi \vdash^+ \psi$

Proof: '$\Rightarrow$': Suppose that $W\Phi \vdash^- \psi$. Since $\vdash^+$ extends $\vdash^-$ we then also have $W\Phi \vdash^+ \psi$. Moreover, we also have that $\Phi \vdash^+ W\Phi$, (meaning that $\Phi \vdash^+ X\varphi$ for any $X \in W, \varphi \in \Phi$). Combining, $\Phi \vdash^+ W\phi$ and $W\Phi \vdash^+ \psi$ then yields $\Phi \vdash^+ \psi$.

'$\Leftarrow$': Suppose that $\Phi \vdash^+ \psi$, with $\pi^+ = \langle \psi_1 \dots \psi_k (= \psi) \rangle$ a $\vdash^+$-proof of ψ from Φ. Let $\wp(\pi)$ be the number of applications of the Necessitation rule in π to non-$\mathbf{S5}_m(G)$-theorems. We prove, for all n and all ψ:

$$\text{If } \Phi \vdash^+ \psi \text{ has a proof } \pi^+ \text{ with } \wp(\pi^+) = n, \text{ then } W\Phi \vdash^- \psi \tag{7}$$

- If $\wp(\pi^+) = 0$, we replace every step in π^+ of the form φ_j, which is justified by $\varphi_j \in \Phi$, by the three steps $K_i\varphi_j, K_i\varphi_j \to \varphi_j, \varphi_j$ to obtain a proof $\kappa(\pi^+)^-$ for $W\Phi \vdash^- \varphi$. (Note that the resulting proof $\kappa(\pi^+)^-$ indeed only uses premises from $W\Phi$.)

- Suppose that (7) holds for al π^+ with $\wp(\pi^+) \leq k$, and we have a $\Phi \vdash^- \psi$ with a proof π^+ with $\wp(\pi^+) = k{+}1$, say $\pi^+ = \langle \psi_1 \dots \psi_r, \psi_{r+1}, \dots \psi_{k+1} (= \psi) \rangle$, where ψ_{r+1} is obtained from $\psi_h (h \leq r)$ as a result of the last Necessitation-step to previous formulas. So, $\psi_{r+1} = K_j\psi_h$, for some $j \leq m$. Thus, if π'^+ is the proof for $\Phi \vdash^+ \psi_h$, then $\wp(\pi'^+) \leq k$. The induction hypothesis yields $W\Phi \vdash^- \psi_h$. By Lemma A.1, we obtain $W\Phi \vdash^- K_j\psi$. Note that the remaining part σ^+ of the $\vdash^+$-proof π^+ with $\sigma^+ = \langle (\psi_{k+1} (= \psi) \dots \psi r + 1 (= K_j\psi_h) \rangle$ from Φ does not involve Necessitation to formulas in this sequence, so $\wp(\sigma^+) = 0$. Thus, as above, we can transform it in a $\vdash^-$-proof $\kappa(\sigma^+)^-)$ for $K_j\psi$ from $W\Phi$ (replacing each assumption in σ^+ of the form $\psi_v (\in \Phi)$ by the triple $K_1\psi_v (\in W\Phi), K_1\psi_v \to \psi_v, \psi_v)$.

■

Proof of Theorem 2.8

- This is a standard result in modal logic, see for instance [9] or [2, 14] for this specific multi-agent epistemic logic $\mathbf{S5}_m(G)$. This property is also referred to as 'weak soundness and completeness'.

- This is the standard notion of adding premises in modal logic. Proofs of this 'strong soundness and completeness' result can again be found in [9] or [2, 14].

- For the $\Rightarrow$-direction, one only has to check that applying Necessitation to premises in Φ is $\models^+$-valid. The argument for $\Leftarrow$ runs as follows. Suppose that we have Φ and φ for which $\Phi \not\vdash^+ \varphi$. By Theorem 2.5, we then have $W\Phi \not\vdash^- \varphi$, where W is defined as in the proof for Theorem 2.5 above. Thus, we find some Kripke model $\mathcal{M}$ and worlds s for which $\mathcal{M}, s \models W\Phi$, $\mathcal{M}, s \not\models \varphi$. Using a standard argument in modal logic (see for instance [13, Theorem 2.12]), one may assume that the model $\mathcal{M}$ is *generated* from s: for all $w \in S$ there is a path $\vec{v} = \langle s = w_1, w_2, \ldots w_n = w\rangle$ such that for each w_i and $w_{i+1}(1 \leq i < n)$ there is some $v_i \leq m$ with $R_{v_i} w_i w_{i+1}$. We now claim that $\mathcal{M} \models \Phi$, $\mathcal{M} \not\models \varphi$, which is equivalent to $\Phi \not\models^+ \varphi$. To show that $\mathcal{M} \models \Phi$, let $\psi \in \Phi$ and let w be an arbitrary element in the set of states, S. Let Since $\mathcal{M}, s \models W\Phi$. Let the path $\vec{v}$ be as above, and similarly for $v_i \leq m$, $R_{v_i} w_i w_{i+1}$. Now, $K_{v_1} K_{v_2} \ldots K_{v_{n-1}} \psi$ is an element of $W\Phi$, so $\mathcal{M}, s \models K_{v_1} K_{v_2} \ldots K_{v_{n-1}} \psi$. But then, also $\mathcal{M}, w \models \psi$, which was to be proven.

■

On Observation 4.4, 4.5

Table 1 should be read as follows: the first column gives the name of the corresponding horizontal line; the second column give possible tuple Λ and then, the 'yes' column give a reason why this tuple gives rise to the claim that the corresponding notion of group knowledge indeed may be called distributed (see the proof of Observation 4.4 below). The 'no' column, to the contrary, argues why the notion cannot be called distributed (see the proof of Observation 4.5). Then, the same exercise is done for the tuple Λ' that is obtained from Λ by changing κ_k from 0 to 1.

Proof of Observation 4.4

We claim that the following tuples $\Lambda = \langle \kappa_k, K, \kappa_l, L, \kappa_g, G\rangle$ make the notion of group knowledge distributed —but in a trivial way.

a Tuples of type $\langle 0, K, 0, L, 0, G\rangle$, i.e., no premises allowed. This was already observed in Observation 4.2, and the argument is given also in Section 4.1.

b Suppose that $\kappa_g = \kappa_l$, and that $G = L$ or $G = \Leftrightarrow, L = +$. Then, if $\kappa_g \cdot \Phi \vdash^G G\varphi$, then also $\kappa_g \cdot \Phi \vdash^G \varphi$ and, under the specified conditions, also $\kappa_l \cdot \Phi \vdash^L (\top \wedge \cdots \top) \rightarrow \varphi$. Of course, one also has $\kappa_k \vdash^K (K_1\top \wedge \cdots \wedge K_m\top)$, independent of κ_k and K. Note how this item generalizes (the proof of) Observation 4.3.

c Tuples Λ for which $\kappa_g = 0$ (note how this item generalizes item a). Under this Λ, suppose that $\vdash G\varphi$. Then, by Theorem 3.3, we have $\vdash (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$ and hence $\kappa_k \vdash^K (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$. Also, one has $\kappa_l \vdash^L (\varphi \wedge \cdots \wedge \varphi) \rightarrow \varphi$. Thus, by this item, we have judged half of all the possibilities to obtain distributed group knowledge as trivial.

d Here, $K = +$ and $\kappa_g \Leftrightarrow \kappa_k$. The crucial observation now is that if $\kappa_g \cdot \Phi \vdash^G G\varphi$, then $\kappa_k \cdot \Phi \vdash^K K_i\varphi$. Obviously, we also have $\kappa_l \vdash^L (\varphi \wedge \cdots \wedge \varphi) \rightarrow \varphi$.

Proof of Observation 4.5

The following tuples $\Lambda = \langle \kappa_k, K, \kappa_l, L, \kappa_g, G \rangle$ make the notion of group knowledge to be not distributed.

e $\kappa_k = \kappa_l = 0$, $\kappa_g = 1$. Let $\Phi = \{Gp\}$. Suppose that we would have $\varphi_1, \ldots, \varphi_m$ for which $\vdash (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$ and $\vdash (\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow p$. Then (using $A3$ and $R1$) we would also have $\vdash p$, which we don't.

f $\kappa_g = 1, G = +$. Let $\Phi = \{p, Gp \rightarrow q\}$. We have $\Phi \vdash^+ Gq$. As a first subcase, suppose that $\kappa_k = 0$ and $L = \Leftrightarrow$. Suppose for this subcase that there are $\varphi_1, \ldots \varphi_m$ with $\vdash^K (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$ and $\{p, Gp \rightarrow q\} \vdash^- (\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow q$. Then we would have $\{p, Gp \rightarrow q\} \vdash^- q$. But the fact that $\{p, Gp \rightarrow q\} \not\models^- q$ is easily verified, and then Theorem 2.8 yields $\{p, Gp \rightarrow q\} \not\vdash^- q$.

g As a second subcase to the previous item, suppose, on top of $\kappa_g = 1, G = +$, that $\kappa_k = 1, K = \Leftrightarrow$ and that either $\kappa_l = 0$ or $L = \Leftrightarrow$. As in the previous item, we then have that if there would be $\varphi_1, \ldots \varphi_m$ for which $\{p, Gp \rightarrow q\} \vdash^- (K_1\varphi_1 \wedge \cdots \wedge K_m\varphi_m)$, and $\kappa_l \cdot \vdash^K (\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow \varphi$, then we would also have $\{p, Gp \rightarrow q\} \not\vdash^- q$, which we don't, as argued above.

h The remaining tuples are $\Lambda = \langle \kappa_k, K, \kappa_l, L, \kappa_g, G \rangle = \langle 1, \Leftrightarrow, 0, L, 1, \Leftrightarrow \rangle$, with $L = 0, 1$. That they do not guarantee distributivity of group knowledge follows from Observation 4.6.

Proof of Observation 4.6

In Observation 4.6 we claimed the following: Although we have $\{K_1 p \vee K_2 p\} \vdash^- Gp$, there exist no φ_1, φ_2 with $\vdash (\varphi_1 \wedge \varphi_2) \to p$ and $\{K_1 p \vee K_2 p\} \vdash^- (K_1\varphi_1 \wedge K_2\varphi_2)$. Consider the model of Figure 2 (reflexive arrows are omitted; R_1 is denoted with the solid line, R_2 with a dashed one).

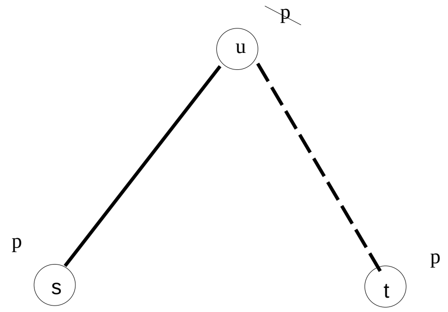

Figure 2: The model proving Observation 4.6

Suppose that there would be formulas φ_1 and φ_2 of the type described above. Since we both have $\mathcal{M}, s \models K_1 p \vee K_2 p$ and $\mathcal{M}, t \models K_1 p \wedge K_2 p$, we have $\mathcal{M}, s \models (K_1\varphi_1 \wedge K_2\varphi_2)$ and $\mathcal{M}, t \models (K_1\varphi_1 \wedge K_2\varphi_2)$. Thus, we have $\mathcal{M}, u \models \varphi_1 \wedge \varphi_2$. However, this is not compatible with $\vdash (\varphi_1 \wedge \varphi_2) \to p$, thus such φ_1 and φ_2 do not exist. ∎

Proof of Corollary 4.7

For $\Phi = \{K_1 p \vee K_2 p\}$ we have $\Phi \vdash^- Gp$, but we also saw that there are no φ_1 and φ_2 such that $\vdash (\varphi_1 \wedge \varphi_2) \to p$ and for which $\Phi \vdash^- (K_1\varphi_1 \wedge K_2\varphi_2)$, which is sufficient to disprove distributivity for the tuples $\Lambda = \langle 1, \not\Leftrightarrow, 0, L, 1, \not\Leftrightarrow \rangle$, $L = +, \not\Leftrightarrow$. ∎

Proof of Observation 4.10

Observation 4.10 claimed about 1 that:

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

Proof: We first state and prove the following lemma:

Lemma A.2 *For all* $u \in \{x, y, z\}$, *and all* φ: $\mathcal{M}, u_1 \models \varphi \Leftrightarrow \mathcal{M}, u_2 \models \varphi$.

Proof: (of Lemma A.2) Using induction on φ. For atoms in Π this follows by the definition of the valuation π. The proof for the connectives $\wedge, \vee$ and $\neg$ is straightforward. So let us assume the statement has been proven for formulas ψ and consider $\varphi = K_1\psi$. If $\mathcal{M}, x_1 \models K_1\psi$, we have that $\mathcal{M}, u_1 \models \psi$ for $u = x, y, z$. The induction hypothesis then yields that $\mathcal{M}, u_2 \models \psi$ for $u = x, y, z$, and thus $\mathcal{M}, x_2 \models K_1\psi$. Similar arguments hold for the other direction, as well as for the y and z worlds in the model, and also for $\varphi = K_2\psi, G\psi$. End of Proof of Lemma A.2 ∎

Now, the first item of Observation 4.10 is proven as follows. Suppose that $\psi \in K(1, \mathcal{M}, x_1)$ that, is, $\mathcal{M}, x_1 \models K_1\psi$. Thus, by definition of K_1, we have $\mathcal{M}, u_1 \models \psi$ for $u = x, y, z$. Then, by the previous lemma, we also have that x_2 and y_2 verify ψ. But then, obviously, we have $\mathcal{M}, x_2 \models K_2\psi$.
To see the second item, observe that $(x_1, v) \in R_1 \cap R_2 \Leftrightarrow (v = x_1 \text{ or } v = y_1)$. And, since q is true in both x_1 and y_1, we have $\mathcal{M}, x_1 \models Gq$.
Finally, since q is false in z_1, we have that K_1q is false at x_1, so $q \notin K(1, \mathcal{M}, x_1)$. The equality $K(1, \mathcal{M}, x_1) \cup K(2, \mathcal{M}, x_1) = K(1, \mathcal{M}, x_2)$ follows immediately from the first item. ∎

Proof of Lemma 5.2

This is easily first obtained for singletons $X = \{x\}$. Since $\mathcal{M} = \langle \mathcal{S}, \ldots \rangle$ is finite, we can enumerate the worlds that differ from x as $s_1, s_2, \ldots s_n$. Let α_x be $(\varphi_{x^+, s_1^-} \wedge \varphi_{x^+, s_2^-} \wedge \ldots \wedge \varphi_{x^+, s_n^-})$. Then, obviously $\mathcal{M}, s \models \alpha_x \Leftrightarrow s = x$. Subsequently, one puts $\alpha_X = \bigvee_{x \in X} \alpha_x$. ∎

Proof of Lemma 5.3

Let $t \in R_i(s)$. We have to show that $\mathcal{M}, t \models \neg\alpha_{X_1} \rightarrow (\neg\alpha_{X_2} \rightarrow (\ldots(\neg\alpha_{X_n} \rightarrow \zeta)\ldots))$ (*). Note that, for arbitrary $x \in \mathcal{S}$ and formulas β, we have $(\mathcal{M}, x \models \neg\alpha_{X_j} \rightarrow \beta) \Leftrightarrow (x \notin X_j \Rightarrow \mathcal{M}, x \models \beta)$. So, (*) is equal to $t \notin X_1 \Rightarrow (t \notin X_2 \Rightarrow (\ldots(t \notin X_n \Rightarrow \mathcal{M}, t \models \zeta)\ldots))$ (**). Now, (**) is true because $t \in (X_1 \cup \ldots \cup X_n \cup Z)$ and, for all $x \in \mathcal{S}$, $x \in Z \Rightarrow \mathcal{M}, x \models \zeta$. ∎

Proof of Theorem 5.4

For the '$\Leftarrow$' direction, suppose $\bigcup_i K(i, \mathcal{M}, s) \models^- \varphi$, and let t be such that $(s, t) \in R_1 \cap \ldots R_m$. Then, for every $\beta \in \bigcup_i K(i, \mathcal{M}, s) \models^- \varphi$ we have $\mathcal{M}, t \models \beta$ and thus, $\mathcal{M}, t \models \varphi$. By definition of t, we may conclude that $\mathcal{M}, s \models G\varphi$.
For the '$\Rightarrow$'-direction, suppose $\mathcal{M}, s \models G\varphi$. Let us consider $R_1(s)$: we can cover it with possibly overlapping regions: let $Z \subseteq R_1(s)$ be $R_1 \cap \ldots \cap R_m$,

and for every agent i, let $X_i = \{x \in R_1(s) \mid x \notin R_i(s)\}$. Thus, X_i is the set of worlds considered possible by agent 1, but not by agent i. If the agents were able to communicate, agent i could tell agent 1 that each $x \in X_i$ is not a good epistemic alternative, so that the agent 1 can eliminate all worlds x that appear in some set X_i, thus ending up in exactly those alternatives in $R_1 \cap \ldots \cap R_m$. Formally, we put Lemma 5.3 to work. Note that $R_1(s) \subseteq (X_1 \cup \ldots \cup X_m \cup Z)$. By Lemma 5.2, we find a characteristic formula α_{X_i} for every X_i, and, by definition of Z, we have $\mathcal{M}, z \models \varphi$, for all $z \in Z$. Then, Lemma 5.3 guarantees that $\mathcal{M}, s \models K_1(\neg\alpha_{X_1} \rightarrow (\neg\alpha_{X_2} \rightarrow (\ldots(\neg\alpha_{X_n} \rightarrow \varphi)\ldots)))$ and, by definition of X_i and α_{X_i}, we also have $\mathcal{M}, s \models K_i\neg\alpha_{X_i}$ (for, if $\mathcal{M}, s \not\models K_i\neg\alpha_{X_i}$ there would be a $t \in R_i(s)$ with $\mathcal{M}, t \models \alpha_{X_i}$, i.e., $t \in X_i$, which contradicts with the definition of X_i). Thus, $\{\neg\alpha_{X_1} \rightarrow (\neg\alpha_{X_2} \rightarrow (\ldots(\neg\alpha_{X_n} \rightarrow \varphi)\ldots), \neg\alpha_{X_1}, \neg\alpha_{X_2}, \ldots \neg\alpha_{X_m}\} \subseteq \bigcup_i K(i, \mathcal{M}, s)$, and thus $\bigcup_i K(i, \mathcal{M}, s) \models^{-} \varphi$. ■

Proof of Observation 5.5

- The proof of this claim is provided by the model of Figure 1
- Take the canonical model $\mathcal{M}^c$ for $\mathbf{S5}_m(G)$. There is a set , $= \{\psi \mid \mathcal{M}, s \models \psi\}$ which is a world in this canonical model $\mathcal{M}^c$, where $\mathcal{M}, s$ is as in Figure 1. , is a maximal consistent set, and hence, a member of the canonical model $\mathcal{M}^c$. Also, one has $\mathcal{M}^c, , \models (K_1\alpha \leftrightarrow K_2\alpha)$, for all α, $\mathcal{M}^c, , \models \neg K_1 p$ and also $\mathcal{M}^c, , \models Gp$. Thus, $\mathcal{M}^c, ,$ demonstrates that the principle of full communication is not violated in the canonical model. This model $\mathcal{M}^c$ moreover is infinite, and distinguishing, by definition (its set of states is the set of all maximal consistent sets). However, we have to mention here that $\mathcal{M}^c$ is in general not a model of the kind as defined in Definition 2.6; the truth definition of $\mathcal{M}, s \models G\varphi$ is $\forall t((s,t) \in R_G \Rightarrow \mathcal{M}, t \models \varphi$, where R_G can only be guaranteed to be some subset of $R_1 \cap \ldots \cap R_m$. For more on this, the reader is referred to [7].

	κ_k	K	κ_l	L	κ_g	G	yes	no	κ_k	K	κ_l	L	κ_g	G	yes	no
1	0	−	0	−	0	−	a,b,c		1	−	0	−	0	−	b,c	
2	0	−	0	−	0	+	a,c		1	−	0	−	0	+	c	
3	0	−	0	−	1	−		e	1	−	0	−	1	−		h
4	0	−	0	−	1	+		e,f	1	−	0	−	1	+		g
5	0	−	0	+	0	−	a,b,c		1	−	0	+	0	−	b,c	
6	0	−	0	+	0	+	a,b,c		1	−	0	+	0	+	b,c	
7	0	−	0	+	1	−		e	1	−	0	+	1	−		h
8	0	−	0	+	1	+		e	1	−	1	+	1	+		g
9	0	−	1	−	0	−	c		1	−	1	−	0	−	c	
10	0	−	1	−	0	+	c		1	−	1	−	0	+	c	
11	0	−	1	−	1	−	b		1	−	1	−	1	−	b	
12	0	−	1	−	1	+		f	1	−	1	−	1	+		g
13	0	−	1	+	0	−	c		1	−	1	+	0	−	c	
14	0	−	1	+	0	+	c		1	−	1	+	0	+	c	
15	0	−	1	+	1	−	b		1	−	1	+	1	−	b	
16	0	−	1	+	1	+	b		1	+	0	+	1	+	b	
17	0	+	0	−	0	−	a,b,c,d		1	+	0	−	0	−	b,c	
18	0	+	0	−	0	+	a,c,d		1	+	0	−	0	+	c	
19	0	+	0	−	1	−		e	1	+	0	−	1	−	d	
20	0	+	0	−	1	+		e,f	1	+	0	−	1	+	d	
21	0	+	0	+	0	−	a,b,c,d		1	+	0	+	0	−	b,c	
22	0	+	0	+	0	+	a,b,c,d		1	+	0	+	0	+	b,c	
23	0	+	0	+	1	−		e	1	+	0	+	1	−	d	
24	0	+	0	+	1	+		e	1	+	1	+	1	+	d	
25	0	+	1	−	0	−	c,d		1	+	1	−	0	−	c	
26	0	+	1	−	0	+	c,d		1	+	1	−	0	+	c	
27	0	+	1	−	1	−	b		1	+	1	−	1	−	b,d	
28	0	+	1	−	1	+		f	1	+	1	−	1	+	d	
29	0	+	1	+	0	−	c,d		1	+	1	+	0	−	c	
30	0	+	1	+	0	+	c,d		1	+	1	+	0	+	c	
31	0	+	1	+	1	−	b		1	+	1	+	1	−	b,d	
32	0	+	1	+	1	+	b		1	+	1	+	1	+	b,d	

Table 1: The table summarizing observations 4.4 and 4.5.