Bisimulation and Distributed Knowledge Revisited

Floris Roelofsen*
Institute for Logic, Language, and Computation
Amsterdam, Netherlands

March 14, 2005

DRAFT

1 Introduction

An important notion in epistemic logic is that of *distributed knowledge* among a group of agents, i.e. the knowledge that is obtained when the individual knowledge of all the agents in the group is put together [1, 4].

The standard semantical definition of distributed knowledge, however, appears to be rather problematic. One major issue is that it does not fit well with the notion of bisimulation [6, 3, 5]. Given the apparent robustness of bisimulation as a notion of epistemic model equivalence this observation can be taken as an argument against the standard semantical account of distributed knowledge [3, 5].

In section 4 however, we will argue that bisimulation is not as suitable a notion of epistemic equivalence as is generally assumed, especially if *groups of agents* are involved. We propose a straightforward generalization of bisimulation, called *collective bisimulation*, which indeed avoids the problem concerning distributed knowledge in a natural way.

Another issue that is sometimes considered as an indication of the unsuitability of the standard semantical definition of distributed knowledge is that it does not fully correspond with an alternative, but at least equally well motivated, more syntactical notion of distributed knowledge [2, 7, 5].

*The ideas presented here have largely benefited from discussions with Johan van Benthem, Michael Franke, Jelle Gerbrandy, Paul Harrenstijn, Fenrong Liu, Maricarmen Martinez, Siewert van Otterloo, Olivier Roy, and Yanjing Wang.

Gerbrandy [2] and van der Hoek et.al. [7] identified special classes of epistemic models on which the semantical and syntactical notions of distributed knowledge do coincide. However, no reasons whatsoever have been recognized as to why one would like to restrict the present semantical framework to either one of these special classes of models. Roelofsen and Wang [5] gave yet another semantical definition of distributed knowledge, which is based on the standard one, but relativized with respect to the model operation of bisimulation contraction, and which does coincide with the syntactical notion of distributed knowledge. But again, this solution is nothing but a technical trick, deprived of any intuitive motivations whatsoever.

Here, we take a different stance. In sections 6 and 7, we will not only explain the difference (and partial overlap) between the standard semantical and the alternative syntactical notion of distributed knowledge, but also argue that it is actually *desirable* to have a plurality of accounts of distributed knowledge. Under certain natural assumptions both notions are suitable, and they indeed coincide in this case; under other assumptions only the semantically motivated notion makes proper sense, and yet under different assumptions the syntactically motivated account should be adopted.

2 Epistemic Logic

We assume a countable set of proposition letters $\mathcal{P}$ and a finite set of agents $\mathcal{A}$ to be given throughout our general discussion and clear from the context in particular examples.

Languages. The basic epistemic language consists of all formulas that can be built from proposition letters in $\mathcal{P}$, using conjunction, negation, and a modal operator K_a for every agent $a \in \mathcal{A}$, where $K_a\varphi$ stands for "agent a knows that φ is the case". We denote the basic epistemic language by $\mathcal{L}^K$:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi$$

One standard way to extend the basic epistemic language is to add modal operators D_G for every group of agents $G \subseteq \mathcal{A}$. $D_G\varphi$ stands for "it is distributed knowledge among G that φ is the case", i.e., φ is a logical consequence of the combined knowledge of all the agents in G. We denote the resulting language by $\mathcal{L}^D$:

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \mid D_G\varphi$$

For any formula $\varphi \in \mathcal{L}^D$, let $d(\varphi)$ denote the maximum number of nested D_G operators in φ: $d(p) = 0$ whenever $p \in \mathcal{P}$, $d(\neg\varphi) = d(\varphi)$, $d(\varphi \wedge \psi) = max(d(\varphi), d(\psi))$, $d(K_a\varphi) = d(\varphi)$, and $d(D_G\varphi) = d(\varphi) + 1$. For each $n = 0, 1, 2, \ldots$, we define:

$$\mathcal{L}^{D_n} = \{\varphi \in \mathcal{L}^D \mid d(\varphi) \leq n\}$$

to be the sub-language of $\mathcal{L}^D$ containing only formulas with at most n nested D_G operators. Notice that $\mathcal{L}^K = \mathcal{L}^{D_0}$ and $\mathcal{L}^D = \cup_n \mathcal{L}^{D_n}$.

Models. An *epistemic model* M for $\mathcal{P}$ and $\mathcal{A}$ is a triple:

$$(W, \sim, V)$$

where W is a nonempty, possibly infinite set of worlds, $\sim: \mathcal{A} \to \wp(W \times W)$ assigns to every agent $a \in \mathcal{A}$ a so-called *indistinguishability relation* $\sim_a \subseteq W \times W$, which is an equivalence relation consisting of all pairs of worlds between which a is unable to distinguish, and finally $V : W \to \mathcal{P} \to \{0, 1\}$ assigns to every world a propositional valuation of $\mathcal{P}$. An *epistemic state* s for $\mathcal{P}$ and $\mathcal{A}$ is a pair (M, w) where $M = (W, \sim, V)$ is an epistemic model for $\mathcal{P}$ and $\mathcal{A}$ and $w \in W$. We say that s is based on M. Let **S5** denote the class of all epistemic models for $\mathcal{P}$ and $\mathcal{A}$.

We will often simply say *model* and *state* instead of epistemic model and epistemic state. If (M, w) is an epistemic state, we call w the *actual world*. If $M = (W, \sim, V)$ is an epistemic model, we will often talk about worlds in M when we actually mean worlds in W.

For a group of agents $G \subseteq \mathcal{A}$ we introduce the following abbreviation, which is non-standard, but convenient for our further discussion.

$$\sim_{\cap G} \quad = \quad \bigcap_{a \in G} \sim_a$$

Information states. Let $M = (W, \sim, V)$ be an epistemic model and let w be a world in W. Then the *individual information state* $[M, w]_a$ of an agent $a \in \mathcal{A}$ in (M, w) consists of all epistemic states (M, v) that a is unable to distinguish from (M, w):

$$[M, w]_a = \{(M, v) \mid w \sim_a v\}$$

Similarly, the *collective information state* $[M, w]_G$ of a group of agents $G \subseteq \mathcal{A}$ in (M, w) consists of all epistemic states (M, v) that none of the agents in G is able to distinguish from (M, w):

$$[M, w]_G = \{(M, v) \mid w \sim_{\cap G} v\}$$

Semantics. The satisfaction relation $\vDash$ between epistemic states and formulas in the basic epistemic language $\mathcal{L}^K$ is recursively defined as follows:

$$
\begin{array}{lll}
M, w \vDash p & \text{iff} & w \in V(p) \\
M, w \vDash \neg\varphi & \text{iff} & M, w \nvDash \varphi \\
M, w \vDash \varphi \wedge \psi & \text{iff} & M, w \vDash \varphi \text{ and } M, w \vDash \psi \\
M, w \vDash K_a\varphi & \text{iff} & M, v \vDash \varphi \text{ for all } (M, v) \in [M, w]_a
\end{array}
$$

The intuitive idea behind the K_a clause is that agent a knows φ if and only if φ is true in all worlds that a is unable to distinguish from the actual world.

There appear to be at least two ways to define the semantics of formulas involving distributed knowledge. The first and most standard alternative is a semantically motivated generalization of the above K_a clause:

$$
M, w \vDash D_G\varphi \quad \text{iff} \quad M, v \vDash \varphi \text{ for all } (M, v) \in [M, w]_G \tag{1}
$$

According to this definition, φ is distributed knowledge among a group of agents G if and only if φ is true in every state that none of the agents in G is able distinguish from the actual world.

The other definition is a syntactically motivated formalization of the idea that distributed knowledge among a group of agents G is what can be derived from the combined knowledge of all agents in G. This definition of the semantics of D_G is given in terms of the notions of *entailment* and *collective knowledge sets*. These notions are defined simultaneously with the satisfaction definition itself. First of all, let $Know_a^{D_{(n)}}(M, w)$ and $Know_G^{D_{(n)}}(M, w)$ denote the *individual knowledge set* of an agent $a \in \mathcal{A}$ and the *collective knowledge set* of a group of agents $G \subseteq \mathcal{A}$, respectively, in an epistemic state (M, w) and w.r.t. a language $\mathcal{L}^{D_{(n)}}$:

$$
\begin{array}{rcl}
Know_a^{D_n}(M, w) & = & \left\{\varphi \in \mathcal{L}^{D_n} \mid (M, w) \vDash^\circ K_a\varphi\right\} \\
Know_G^{D_n}(M, w) & = & \bigcup_{a \in G} Know_a^{D_n}(M, w) \\
Know_a^{D}(M, w) & = & \bigcup_{n \in \mathbb{N}} Know_a^{D_n}(M, w) \\
Know_G^{D}(M, w) & = & \bigcup_{a \in G} Know_a^{D}(M, w)
\end{array}
$$

A formula $\phi \in \mathcal{L}^{D_n}$ is entailed by a set of formulas $\Phi \subseteq \mathcal{L}^{D_n}$ if and only if all epistemic states that satisfy Φ (i.e., all formulas in Φ) also satisfy ϕ:

$$
\Phi \Vdash \phi \quad \text{iff} \quad \forall (M, w) : M, w \vDash \Phi \Rightarrow M, w \vDash \phi
$$

Finally, a formula $D_G\phi$, where $\phi \in \mathcal{L}^{D_n}$, is satisfied by an epistemic state (M, w) if and only if ϕ is entailed by the collective knowledge set of G in (M, w) w.r.t. $\mathcal{L}^{D_n}$:

$$(M, w) \vDash D_G\phi \quad (\phi \in \mathcal{L}^{D_n}) \quad \text{iff} \quad Know_G^{D_n}(M, w) \Vdash \phi \tag{2}$$

Notice that, in particular, a formula which does not involve distributed knowledge is distributed among G if and only if it is entailed by the collective knowledge set of G w.r.t. $\mathcal{L}^K$:

$$(M, w) \vDash D_G\phi \quad (\phi \in \mathcal{L}^{K}) \quad \text{iff} \quad Know_G^{K}(M, w) \Vdash \phi$$

For further reference, let us write $\vDash^\bullet$ for the satisfaction relation whose D_G clause is given by 1, and $\vDash^\circ$ for the satisfaction relation whose D_G clause is given by 2.

3 Problems

As mentioned above, $\vDash^\bullet$ is generally accepted as a suitable semantics for distributed knowledge [1, 4]. Also, $\vDash^\bullet$ and $\vDash^\circ$ are often assumed to give an equivalent account of distributed knowledge, as their underlying intuitions really seem to be two formulations of one and the same idea.

But first impressions have turned out to be deceptive in this case: $\vDash^\bullet$ and $\vDash^\circ$ have turned out *not* to be equivalent after all [2, 7, 5]. And $\vDash^\bullet$ itself is problematic for another reason as well: it does not fit well with the notion of bisimulation, which is the standard notion of model equivalence in epistemic logic [6, 3, 5]. In this section we make these issues precise.

3.1 Bisimulation Invariance

Bisimulation is the standard notion of equivalence in epistemic model theory.

Definition 1 (Bisimulation) *Let $M = (W, \sim, V)$ and $M' = (W', \sim', V')$ be two epistemic models. A non-empty binary relation $\equiv\ \subseteq W \times W'$ is a bisimularity relation between M and M' if and only if for every $w \in W$ and $w' \in W'$ such that $w \equiv w'$ we have:*

1. *For every proposition letter $p \in \Phi$:*
   - $V(w)(p) = V'(w')(p)$
2. *For every agent $a \in \mathcal{A}$:*

- *if $w \sim_a v$, then for some $v' \in W'$: $w' \sim'_a v'$ and $v \equiv v'$,*
- *if $w' \sim'_a v'$, then for some $v \in W$: $w \sim_a v$ and $v \equiv v'$.*

We say that two models M and M' are bisimilar, and write $M \simeq M'$, if and only if there is bisimilarity relation between them. We say that two states (M, w) and (M', w') are bisimilar, and wirte $(M, w) \simeq (M', w')$, if and only if there is a bisimularity relation $\equiv$ between M and M' such that $w \equiv w'$. The relation of bisimulation, which is denoted by $\simeq$, is the relation that holds between any two bisimilar states.

Definition 2 (Invariance) *An epistemic language*[1] $\mathcal{L}^{\Delta}$ *is invariant under bisimulation w.r.t. a satisfiability relation* $\vDash$ *if and only if for every two bisimilar states* (M, w) *and* (M', w')*, and for every formula* $\varphi \in \mathcal{L}^{\Delta}$*, we have:*

$$M, w \vDash \varphi \quad \textit{iff} \quad M', w' \vDash \varphi$$

The following proposition[2] has given rise to the generally accepted conception of two bisimilar states as representations of one and the same situation.

Proposition 1 $\mathcal{L}^K$ *is invariant under bisimulation w.r.t. both* $\vDash^{\bullet}$ *and* $\vDash^{\circ}$*.*

Proof. The result for $\vDash^{\bullet}$ is standard [6], and the result for $\vDash^{\circ}$ follows from the fact that $\vDash^{\bullet}$ and $\vDash^{\circ}$ coincide for all formulas in $\mathcal{L}^{KC}$. □

But when the D_G operator for distributed knowledge are added to the language, and when formulas involving distributed knowledge are interpreted according to $\vDash^{\bullet}$, bisimulation is not a suitable notion of model equivalence anymore.

Proposition 2 $\mathcal{L}^D$ *is invariant under bisimulation w.r.t.* $\vDash^{\circ}$*, but not w.r.t.* $\vDash^{\bullet}$*.*

Proof. First consider the positive result. Towards a contradiction, suppose $\mathcal{L}^D$ is not invariant under bisimulation w.r.t. $\vDash^{\circ}$. Then there are two bisimilar states (M, w) and (M', w') and a formula $\varphi \in \mathcal{L}^D$ such that:

$$\begin{array}{lll} M, w & \vDash^{\circ} & \varphi \\ M', w' & \nvDash^{\circ} & \varphi \end{array}$$

[1]Henceforth, $\mathcal{L}^{\Delta}$ is supposed to stand for $\mathcal{L}^K$, $\mathcal{L}^D$, or $\mathcal{L}^{D_n}$ for some $n \in \mathbb{N}$.

[2]together with its successful generalization to the epistemic language with common knowledge operators.

First observe that φ must be in $\mathcal{L}^{D_n}$ for some $n \in \mathbb{N}$. We proof that there must also be a formula $\varphi' \in \mathcal{L}^{D_{n-1}}$ such that $M, w \models^\circ \varphi'$ and $M', w' \not\models^\circ \varphi'$. This implies that there must in fact be a formula $\varphi'' \in \mathcal{L}^K$ with the same properties, which contradicts proposition 1.

We consider the case where φ is of the form $D_G\phi$ for some $\phi \in \mathcal{L}^{D_{n-1}}$. Suppose that $M, w \models^\circ D_G\phi$. Then, by definition, $Know_G^{D_{n-1}}(M, w) \Vdash^\circ \phi$. By compactness, there is a finite subset $\{\psi_1, \ldots, \psi_n\}$ of $Know_G^{D_{n-1}}(M, w)$ such that $\{\psi_1, \ldots, \psi_n\} \Vdash^\circ \phi$ and therefore also $\psi \Vdash^\circ \phi$, where $\psi = \psi_1 \wedge \ldots \wedge \psi_n$. Notice that $\psi \in \mathcal{L}^{D_{n-1}}$. We now show that $M, w \models^\circ \psi$ and $M', w' \not\models^\circ \psi$. For the first, observe that, as everything which is known to be true in (M, w) must in fact be true in (M, w), we have that $M, w \models^\circ \psi_i$ for every $i \in \{1, \ldots, n\}$ and thus that $M, w \models^\circ \psi$. For the second, notice that if $M', w' \models^\circ \psi$ were the case then would have $Know_G^\Delta(M', w') \Vdash^\circ \phi$, and therefore $M', w' \models^\circ D_G\phi$ would also hold, which would contradict our assumption. So $M', w' \not\models^\circ \psi$. This establishes that $\mathcal{L}^D$ is invariant under bisimulation w.r.t. $\models^\circ$.

For the negative result, compare the epistemc states given in figures 1 and 2 (in figures, the solid world always denotes the actual world).

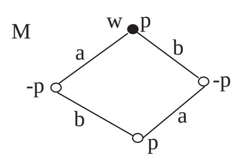


Figure 1: (M, w).

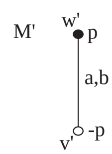


Figure 2: (M', w').

Let $G = \{a, b\}$. Although (M, w) and (M', w') are bisimilar, we have:

$$\begin{array}{rcl}(M, w) & \models^\bullet & D_G p \\ (M', w') & \not\models^\bullet & D_G p\end{array}$$

□

Proposition 2 indicates that $\models^\bullet$, does not fit well with the notion of bisimulation. Given the apparent robustness of bisimulation as a notion of epistemic model equivalence this observation can be taken as an argument against the semantic account of distributed knowledge given by $\models^\bullet$ [3, 5]. In section 4 however, we will argue that bisimulation is not as suitable a notion of epistemic equivalence as is generally assumed if *groups of agents* are involved. We will propose a straightforward generalization of bisimulation, which restores invariance of $\mathcal{L}^D$ w.r.t. $\models^\bullet$ in a natural way.

3.2 Satisfaction

We proceed with a comparison of $\models^{\bullet}$ and $\models^{\circ}$. We first observe that, by definition, $\models^{\bullet}$ and $\models^{\circ}$ coincide for every formula in which D_G does not occur.

Fact 1 *For every formula $\varphi \in \mathcal{L}^K$ and every epistemic state (M, w):*

$$M, w \models^{\circ} \varphi \quad \textit{iff} \quad M, w \models^{\bullet} \varphi$$

As soon as distributed knowledge is involved, however, things are different: $\models^{\bullet}$ turns out to be strictly stronger than $\models^{\circ}$.

Proposition 3 *For every state (M, w) and every formula $\varphi \in \mathcal{L}^D$:*

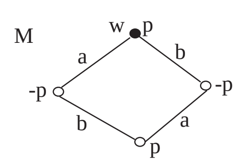
$$M, w \models^{\circ} \varphi \quad \textit{implies} \quad M, w \models^{\bullet} \varphi$$

Proof. It is enough to prove that the statement holds for every formula $\varphi \in \mathcal{L}^{D_n}$, for any $n \in \mathbb{N}$. The proof is by induction on n. The case for $n = 0$ follows directly from fact 1. Next, suppose φ is of the form $D_G\phi$, where $\phi \in \mathcal{L}^{D_n}$, and that $M, w \models^{\circ} D_G\phi$. Then, by definition, $Know_G^{D_n}(M, w) \Vdash^{\circ} \phi$. Now let v be any world in W such that $w \sim_{\cap G} v$. Clearly, for every $\psi \in Know_G^{D_n}(M, w)$ we have $M, v \models^{\circ} \psi$. By the induction hypothesis, $M, v \models^{\circ} \psi$ implies $M, v \models^{\bullet} \psi$, which in turn gives us $M, w \models^{\bullet} D_G\phi$. $\square$

Proposition 4 *There is a state (M, w) and a formula $\varphi \in \mathcal{L}^D$ such that:*

$$\begin{array}{lll} M, w & \not\models^{\circ} & \varphi \\ M, w & \models^{\bullet} & \varphi \end{array}$$

Proof. Recall the epistemic states depicted in figure 1. Let $G = \{a, b\}$.



Every world in M that neither agent a nor agent b is able to distinguish from w (the only such world is w itself) satisfies p. Consequently, $M, w \models^{\bullet} D_Gp$. On the other hand, it is easy to see that $Know_G^K(M, w) \not\Vdash^{\circ} p$, and therefore $M, w \not\models^{\circ} D_Gp$. $\square$

Proposition 4 demonstrates that $\models^\bullet$ and $\models^\circ$ define different semantics for distributed knowledge. Similar results have been established earlier by Gerbrandy [3], van der Hoek et.al. [7] and Roelofsen and Wang [5]. Gerbrandy [2] and van der Hoek et.al. [7] identified special classes of epistemic models on which $\models^\bullet$ and $\models^\circ$ do coincide. However, no motivation is provided as to why one would like to restrict the present semantical framework to either one of these special classes of models. Roelofsen and Wang [5] give an alternative semantics for distributed knowledge, which does coincide with $\models^\circ$. This semantics is based on $\models^\bullet$, but relativized to the model operation of bisimulation contraction.

Here, we take a different approach. In sections 6 and 7, we will not only explain the difference (and partial overlap) between $\models^\bullet$ and $\models^\circ$, but also argue that it is actually *desirable* to have a plurality of semantical accounts of distributed knowledge. Under certain natural assumptions, $\models^\bullet$ and $\models^\circ$ are both suitable, and indeed coincide; under other assumptions only $\models^\bullet$ makes proper sense, and yet under different assumptions only $\models^\circ$ should be considered.

4 Collective Bisimulation

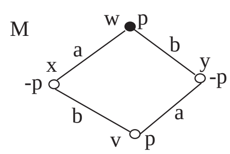
We first consider the problem brought up in section 3.1: that of $\models^\bullet$ not preserving $\mathcal{L}^D$ under bisimulation. Recall that bisimulation is the relation that holds between any two bisimilar epistemic states. Gerbrandy [3] shows that bisimulation can be characterized as the largest equivalence relation between epistemic states that complies with what he calls the principle of epistemic extensionality.

Definition 3 (Principle of Epistemic Extensionality)
An equivalence relation $\equiv$ between epistemic states is called epistemically extensional if it is such that $(M, w) \equiv (M', w')$ if and only if:

1. *For every proposition letter $p \in \Phi$:*
   - $V(w)(p) = V'(w')(p)$
2. *For every agent $a \in \mathcal{A}$:*
   - $[M, w]_a \equiv [M', w']_a$

where $A \equiv B$ for two sets of epistemic states A and B, means that for every state in A we can find an equivalent one in B and vice versa.

We will argue that the principle of extensionality as stated here is too weak to serve as an appropriate foundation for a notion of equivalence between epistemic states. Besides identical valuations of proposition letters and equivalence of *individual* information states, the principle should also enforce equivalence of *collective* information states. The latter should be made explicit, as it does not automatically follow from the former. A case in point is the epistemic state depicted in figure 1:



The individual information state $[M, w]_a$ of agent a in state (M, w) is equivalent to its information state $[M, x]_a$ in (M, x). The same goes for agent b. However, for $G = \{a, b\}$ we have:

$$\begin{aligned} [M, w]_G &= \{(M, w)\} \\ [M, x]_G &= \{(M, x)\} \end{aligned}$$

and (M, w) and (M, x) are in no natural sense equivalent, as they do not satisfy the same proposition letters.

This observation naturally suggests the following adaptation of the principle of epistemic extensionality:

Definition 4 (Revised Principle of Epistemic Extensionality) *An equivalence relation $\equiv$ between epistemic states is called epistemically extensional if it is such that $(M, w) \equiv (M', w')$ if and only if:*

1. *For every proposition letter $p \in \Phi$:*
   - $V(w)(p) = V'(w')(p)$
2. *For every group of agents $G \subseteq \mathcal{A}$:*
   - $[M, w]_G \equiv [M', w']_G$

Notice that the revised principle is a proper generalization of the original principle, which only requires equivalence of information states of all groups of agents consisting of just one agent. So every equivalence relation that complies with the revised principle of epistemic extensionality, also complies with the original principle of epistemic extensionality, but not vice versa.

We call the largest equivalence relation between epistemic states that complies with the revised principle of epistemic extensionality *collective bisimulation*. This relation can also be defined directly as follows.

Definition 5 (Collective Bisimulation) *Let $M = (W, \sim, V)$ and $M' = (W', \sim', V')$ be two epistemic models. A non-empty binary relation $\equiv_c \subseteq W \times W'$ is a collective bisimularity relation between M and M' if and only if for every $w \in W$ and $w' \in W'$ such that $w \equiv_c w'$ we have:*

1. *For every proposition letter $p \in \Phi$:*
   - $V(w)(p) = V'(w')(p)$
2. *For every group of agents $G \subseteq \mathcal{A}$:*
   - *if $w \sim_{\cap G} v$, then for some $v' \in W'$: $w' \sim'_{\cap G} v'$ and $v \equiv_c v'$,*
   - *if $w' \sim'_{\cap G} v'$, then for some $v \in W$: $w \sim_{\cap G} v$ and $v \equiv_c v'$.*

We say that two models M and M' are collectively bisimilar, and write $M \simeq_c M'$, if and only if there is a collective bisimilarity relation between them. We say that two states (M, w) and (M', w') are collectively bisimilar, and wirte $(M, w) \simeq_c (M', w')$, if and only if there is a collective bisimularity relation $\equiv_c$ between M and M' such that $w \equiv_c w'$. The relation of collective bisimulation, which is denoted by $\simeq_c$, is the relation that holds between any two collectively bisimilar states.

As expected, collective bisimulation is a proper generalization of ordinary bisimulation. The latter only requires the "zigzag" conditions to hold for groups consisting of one single agent only. So if two models or states are collectively bisimilar, then they are in any case bisimilar, but not vice versa.

Theorem 1 (Collective Bisimulation Invariance)
$\mathcal{L}^D$ is invariant under collective bisimulation w.r.t. both $\models^\circ$ and $\models^\bullet$.

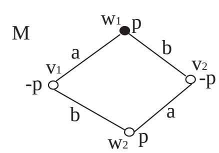
Proof. The result for $\models^\circ$ follows directly from the more general result concerning ordinary bisimulation proven in proposition 2. We proof the claim for $\models^\bullet$, which could not be established when ordinary bisimulation was considered (see proposition 2).

Let $M = (W, \sim, V)$ and $M' = (W', \sim', V')$ be two epistemic models and let $w \in W$ and $w' \in W'$ be such that $(M, w) \simeq_c (M', w')$ are collectively bisimilar. Then we must prove that for all $\varphi \in \mathcal{L}^D$:

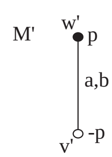
$$M, w \models^\bullet \varphi \quad \text{iff} \quad M', w' \models^\bullet \varphi \tag{3}$$

Clearly, it is enough to prove that the statement holds for every $\varphi \in \mathcal{L}^{D_n}$, for any $n \in \mathbb{N}$. Moreover, it suffices to proof either the "if" or the "only if" part of the statement. We do the latter by induction on n. The case for $n = 0$ follows directly from proposition 1. Now, suppose φ is of the form $D_G\phi$, where $\phi \in \mathcal{L}^{D_n}$. We proceed by contraposition, i.e., we show that $M, w \not\models^{\bullet} D_G\phi$ implies $M', w' \not\models^{\bullet} D_G\phi$. Suppose $M, w \not\models^{\bullet} D_G\phi$. Then $M, v \models^{\bullet} \neg\phi$ for some $v \in W$ such that $w \sim_{\cap G} v$. As $(M, w) \simeq_c (M', w')$, there must be a world $v' \in W'$ such that $w' \sim'_{\cap G} v'$ and $(M, v) \simeq_c (M', v')$. By the induction hyposthesis, then, $M', v' \models^{\bullet} \neg\phi$, and thus $M', w' \not\models^{\bullet} \phi$. □

Theorem 1 brings the good news about collective bisimulation. However, even though it generalizes ordinary bisimulation in a most natural way, collective bisimulation also exhibits an evident and seemingly highly undesirable drawback[3]: it is not *preserved under contraction*[4]. That is, if M is an epistemic model, and $M^{\simeq_c}$ is its collective bisimulation contraction, which is obtained from M by identifying all collectively bisimilar worlds, then it is not generally the case that M and $M^{\simeq_c}$ are collectively bisimilar. The model depicted in figure 1 again serves as a case in point.



Clearly, $(M, w_1) \simeq_c (M, w_2)$ and $(M, v_1) \simeq_c (M, v_2)$. By identifying these worlds we obtain the model depicted in figure 2:



Clearly, M and M' are not collectively bisimilar. Summarizing:

Proposition 5 (Preservation under Contraction)
Collective bisimulation is not preserved under contraction.

[3]This was pointed out to me by Yanjing Wang.

[4]*Preservation under contraction* is called *preservation under quotients*, and argued to be a desirable property of any reasonable notion of epistemic model equivalence by Gerbrandy [3].

In sections 6 and 7, however, we will argue that only under certain assumptions does it make sense to require from a notion of epistemic model equivalence that it be preserved under contraction, and that in fact, under these assumptions, collective bisimulation is indeed preserved under contraction. The assumptions we allude to here are introduced in the next section.

5 Distinguishability

Consider the indistinguishability relation $\sim_a$ associated with some agent $a \in \mathcal{A}$ in a model $M = (W, \sim, V)$. For any two worlds $w, v \in W$, $w \sim_a v$ represents the fact that agent a is not able to distinguish between w and v. This appears to be a univocal intuition, but in fact one crucial aspect remains unspecified here. The following distinction will turn out to be essential for the remainder of our discussion.

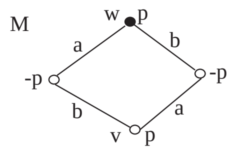
Definition 6 (Strong and General Distinguishability)
Let $M = (W, \sim, V)$ be an epistemic model, let $w, v \in W$, and let $\mathcal{L}^{\Delta}$ be the epistemic language that is used to talk about M.

Strong Distinguishability Assumption *If for every agent a, $w \sim_a v$ is intended to mean that a is able to distinguish w and v in terms of what he knows about expressions in $\mathcal{L}^{\Delta}$ only, then we say that $\sim$ is assumed to represent a strong ability to distinguish worlds. We call this assumption the* SD *assumption.*

General Distinguishability Assumption *If for every agent a, $w \sim_a v$ is intended to mean that a is somehow able to distinguish w and v, possibly in terms of expressions that are not in $\mathcal{L}^{\Delta}$, or even by non-linguistic means (e.g., w and v may "feel" different in some unverbalizable way), then we say that $\sim$ is assumed to represent a general ability to distinguish worlds. We call this the* GD *assumption.*

Under the GD assumumption, i.e., if $\sim$ is assumed to represent a general ability to distinguish worlds, every epistemic model constitutes a reasonable representation of which worlds agents are able to distinguish and which not. However, under the SD assumption, i.e., if $\sim$ is assumed to represent a strong ability to distinguish worlds, some epistemic models do not make proper sense anymore and should be left out of consideration.

Example 1 *Recall the epistemic model depicted in figure 1:*



Notice that w and v are not distinguishable for agent a as far as the part of his knowledge that is expressable in $\mathcal{L}^D$ is concerned:

$$Know_a^D(M,w) = Know_a^D(M,v)$$

The indistinguishability relation $\sim_a$ for a does not reflect this: $w \nsim_a v$.

This example gives rise to the following completeness constraint on the *indistinguishability relations* of epistemic models:

$\sim$completeness Let $M = (W, \sim, V)$ be an epistemic model and let $\mathcal{L}^\Delta$ be an epistemic language. Then M is $\sim$complete w.r.t. $\mathcal{L}^\Delta$ if and only if for every two worlds $w, v \in W$ and every agent $a \in \mathcal{A}$:

$$Know_a^\Delta(M,w) = Know_a^\Delta(M,v) \quad \text{implies} \quad w \sim_a v$$

A similar completeness constraint could be imposed on the *domains* of our models. If in an epistemic state (M, w), the collective knowledge of some group of agents G is consistent with a formula φ, but does not entail it, then M must provide an explicit reason for why this is so, i.e., there must be a world v in M which G is collectively unable to distinguish from w and which satisfies $\neg\varphi$.

w-completeness Let $M = (W, \sim, V)$ be an epistemic model and let $\mathcal{L}^\Delta$ be an epistemic language. Then M is w-complete w.r.t. $\mathcal{L}^\Delta$ if and only if for every world $w \in W$, every group of agents $G \subseteq \mathcal{A}$, and every formula $\varphi \in \mathcal{L}^\Delta$ such that:

$$Know_G^\Delta(M,w) \not\models^\circ \varphi \quad \text{and} \quad \{\varphi\} \cup Know_G^\Delta(M,w) \text{ is } \models^\circ\text{-satisfiable.}$$

there is a world $v \in W$ such that[5]:

$$w \sim_{\cap G} v \quad \text{and} \quad M, v \models^\circ \neg\varphi$$

[5]CLARIFY why we use of $\models^\circ$ rather than $\models^\bullet$ and consider only knowledge in languages without D_G.

Together, $\sim$completeness and w-completeness characterize the class of what we call *acceptable* epistemic models under the SD assumption.

Definition 7 (Acceptable Models) *Let M be an epistemic model and let $\mathcal{L}^{\Delta}$ be an epistemic language. Then we call M acceptable w.r.t. $\mathcal{L}^{\Delta}$ if and only if it is both $\sim$complete and w-complete w.r.t. $\mathcal{L}^{\Delta}$. Let $\boldsymbol{S5}^{\Delta}$ denote the class of all epistemic models acceptable w.r.t. $\mathcal{L}^{\Delta}$.*

We will proceed to show that the distinction between modeling a strong and a general ability of agents to distinguish possible worlds, together with the observation that under the SD assumption certain models should be left out of consideration, clarifies the non-preservation of collective bisimulation under contraction as well as the difference between $\models^{\bullet}$ and $\models^{\circ}$ discussed in section 3.2.

6 Modeling Strong Distinguishability

Let us assume that $\sim$ represents a strong ability to distinguish worlds. As argued in the previous section, under this assumption it makes sense to restrict ourselves to acceptable models. This restriction restores the desired preservation of collective bisimulation under contraction, as well as the harmony between $\models^{\bullet}$ and $\models^{\circ}$. We first consider the preservation result.

Theorem 2 (Preservation under Contraction) *Every epistemic model M that is acceptable w.r.t. some epistemic language $\mathcal{L}^{\Delta}$ is collectively bisimilar to its collective bisimulation contraction: $M \simeq_c M^{\simeq_c}$.*

Proof. It is enough to proof that for every two worlds w and v in M:

$$(M, w) \simeq_c (M, v) \quad \text{implies} \quad w \sim_a v \quad \text{for all } a \in \mathcal{A}$$

which ensures that w and v are identified in $M^{\simeq_c}$ only if no agent (and therefore no group of agents) is able to distinguish one from the other.

Let M be a model acceptable w.r.t. $\mathcal{L}^{\Delta}$, and let w and v be two worlds in M such that $(M, w) \simeq_c (M, v)$. By proposition 1, (M, w) and (M, v) satisfy exactly the same formulas in $\mathcal{L}^{\Delta}$. In particular, $Know_a^{\Delta}(M, w) = Know_a^{\Delta}(M, v)$ for every agent $a \in \mathcal{A}$. But then, by $\sim$completeness of M w.r.t. $\mathcal{L}^{\Delta}$, we may conclude that $w \sim_a v$ for every agent $a \in \mathcal{A}$. □

Next, we show that $\models^{\bullet}$ and $\models^{\circ}$ coincide on acceptable models.

Theorem 3 (Satisfaction) *Let $\mathcal{L}^{\Delta}$ be an epistemic language, and let M be an acceptable model w.r.t. $\mathcal{L}^{\Delta}$. Then, for every formula $\varphi \in \mathcal{L}^{\Delta}$ and every world w in M:*

$$M, w \models^{\circ} \varphi \quad \textit{iff} \quad M, w \models^{\bullet} \varphi$$

Proof. The result is trivial for $\mathcal{L}^K$, as per definition, $\models^{\circ}$ and $\models^{\bullet}$ coincide for all formulas in $\mathcal{L}^K$. The "only if" part of the statement for $\mathcal{L}^D$ follows directly from the more general result in proposition 3. To estabish the "if" part for $\mathcal{L}^D$ it is sufficient to proof the corresponiding statement for $\mathcal{L}^{D_n}$, $n \in \mathbb{N}$. This can be done by induction on n. The case for $n = 0$ follows from the above remarks regarding $\mathcal{L}^K$. Next, suppose φ is of the form $D_G\phi$, where $\phi \in \mathcal{L}^{D_n}$. We proceed by contraposition. That is, we assume $M, w \not\models^{\circ} D_G\phi$ and show that $M, w \not\models^{\bullet} D_G\phi$. If $M, w \not\models^{\circ} D_G\phi$, then, by definition, $Know_G^{D_n}(M, w) \not\Vdash^{\circ} \phi$, which means that $\{\neg\phi\} \cup Know_G^{D_n}(M, w)$ is $\models^{\circ}$-satisfiable. Now consider two cases. (1) If $Know_G^{D_n}(M, w) \not\Vdash^{\circ} \neg\phi$, then, by w-completeness of M, there must be a world v in M such that $w \sim_{\cap G} v$ and $M, v \models^{\circ} \neg\phi$. From the induction hypothesis it follows that $M, v \models^{\bullet} \neg\phi$, which yields $M, w \not\models^{\bullet} D_G\phi$, as desired. (2) If $Know_G^{D_n}(M, w) \Vdash^{\circ} \neg\phi$, then $M, w \models^{\circ} \neg\phi$ follows directly. Again, by the induction hypothesis, we get $M, w \models^{\bullet} \neg\phi$ and thus $M, w \not\models^{\bullet} D_G\phi$, as desired. □

7 Modeling General Distinguishability

Now let us assume that $\sim$ represents a general ability to distinguish worlds. Then all epistemic models make proper sense. If, according to $\sim$, two worlds in M are distinguishable for an agent a, then they are so *for a reason*, even if this reason may not be expressable in the epistemic language that is used to talk about the model. Within this perspective, identifying collectively bisimilar worlds, or any worlds whatsoever, does not make sense. So the problem of collective bisimulation not being preserved under contraction, brought up in section 4, becomes a non-issue under the GD assumption.

Now consider the difference between $\models^{\bullet}$ and $\models^{\circ}$. The intuitive idea behind $\models^{\bullet}$ is that a formula φ is distributed knowledge among a group of agents G if and only if φ holds in every world that neither of the agents in G is able to distinguish from the actual world.

This idea only makes sense if we assume that the distinctions which all the individual agents are able to make between different possible worlds can somehow be combined into an overall distinction between worlds that they are able to make collectively. In other words, we must assume that the agents are able to *communicate* with each other, in some language they all understand, about which worlds they are able to distinguish. Furthermore, we must assume that the language used by the agents to do so is generally more expressive than the epistemic language that is used to talk about the model, given that their ability to distinguish different possible worlds may not to be fully expressable in the latter epistemic language. For further reference, let us make a clear distinction between this assumption and its more restrictive counterpart.

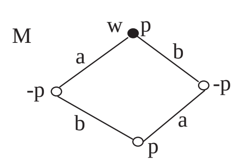
Definition 8 (Communication Language Assumptions) *Let $\mathcal{L}_{COM}$ be the language that the agents use to communicate about which worlds they are able to distinguish, and let $\mathcal{L}^{\Delta}$ be the epistemic language that is used to talk about the model of the agents' information.*

Restricted communication language assumption
*We call the assumption that $\mathcal{L}_{COM} = \mathcal{L}^{\Delta}$ the restricted communication language assumption (*RCL*).*

Unrestricted communication language assumption
*We call the assumption that $\mathcal{L}_{COM}$ is expressive enough for all the agents to communicate which worlds they are able to distinguish the unrestricted communication language assumption (*UCL*).*

Under the RCL assumption, $\models^{\circ}$ makes proper sense: according to this satisfaction relation, distributed knowledge of G is exactly what follows from the combined knowledge of the agents in G, as far as this knowledge is expressable in the modeling language $\mathcal{L}^{\Delta}$.

On the other hand, $\models^{\bullet}$ should be rejected under the RCL assumption. To see this, consider the model from figure 1 which is depicted again below:



In terms of $\mathcal{L}^D$ the agents are not able to distinguish any world in (M, w) from any other: the agents' knowledge in terms of $\mathcal{L}^D$ is exactly the same in all worlds. So if the agents' communication language is limited to $\mathcal{L}^D$, then it is not fair to conclude, as $\models^{\bullet}$ does, that p is distributed knowledge among a and b in (M, w).

Under the UCL assumption, tables are turned: $\models^{\circ}$ is too weak, whereas $\models^{\bullet}$ neatly fits the intuitions. In the model depicted above, p should now be considered distributed knowledge among a and b in (M, w), which $\models^{\bullet}$ confirms, but $\models^{\circ}$ erroneously denies.

We conclude that under the RCL assumption $\models^{\circ}$ is the proper satisfaction relation to use, whereas under the UCL assumption $\models^{\bullet}$ should be adopted. In this light, the difference between $\models^{\circ}$ and $\models^{\bullet}$ is problematic at all. On the contrary, each semantics suits a different assumption about the agents' communication language.

8 Discussion

SD	GD
Collective bisimulation preserved under contraction.	Preservation under contraction irrelevant.
$\models^{\circ}$ and $\models^{\bullet}$ coincide.	Use $\models^{\circ}$ under RCL assumption; Use $\models^{\bullet}$ under UCL assumption.

The above diagram summarizes our observations. We argued that bisimulation is too weak a notion of equivalence between epistemic states when (the distributed knowledge among) a group of agents is involved. We introduced the notion of collective bisimulation as a natural generalization of its traditional counterpart, and showed that it is indeed a more suitable notion of equivalence between collective information states. An apparent drawback of collective bisimulation (the fact that it is not generally preserved under contraction) was shown to dissolve under the SD assumption (i.e., when $\sim$ is assumed to represent strong distinguishability) and to be irrelevant under the GD assumption (i.e., when $\sim$ is assumed to represent general distinguishability).

We also discussed the difference between $\models^{\circ}$, which models distributed knowledge as what follows from the combined knowledge of a group of agents, and $\models^{\bullet}$, which models distributed knowledge as what holds in all worlds that neither agent is able to distinguish from the actual world. We showed that $\models^{\circ}$ and $\models^{\bullet}$ coincide under the SD assumption, and argued that the contrast between them under the GD assumption is actually desired. If the agents' communication language is assumed to be restricted to the epistemic language that is used to talk about the model, then $\models^{\circ}$ should be adopted. If not, $\models^{\bullet}$ provides a more sensible semantics.

References

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

[2] J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, University of Amsterdam, 1998.

[3] J. Gerbrandy. Epistemic equivalence and bisimulation. Manuscript. 2005.

[4] J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Campbridge University Press, 1995.

[5] F. Roelofsen and Y. Wang. Distributed knowledge and bisimulation contraction. Manuscript. 2005.

[6] J. van Benthem. One is a lonely number: on the logic of communication. *Technical Report PP-2002-27, ILLC Amsterdam*, 2002.

[7] W. van der Hoek, B. van Linder, and J. Meyer. Group knowledge is not always distributed (neither is it always implicit). *Mathematical social sciences*, 38:215–240, 1999.