Common knowledge logic in a higher order proof assistant?

Pierre Lescanne

Common knowledge logic in a higher order proof assistant?

Pierre Lescanne Laboratoire de l'Informatique du Parallélisme,
École Normale Supérieure de Lyon
46, Allée d'Italie, 69364 Lyon 07, FRANCE
Pierre.Lescanne@ens-lyon.fr

December 19, 2007

Abstract

This paper presents experiments on common knowledge logic, conducted with the help of the proof assistant Coq. The main feature of common knowledge logic is the eponymous modality that says that a group of agents shares a knowledge about a certain proposition in a inductive way. This modality is specified by using a fixpoint approach. Furthermore, from these experiments, we discuss and compare the structure of theorems that can be proved in specific theories that use common knowledge logic. Those structures manifests the interplay between the theory (as implemented in the proof assistant Coq) and the metatheory.

1 Introduction

In a previous paper [13], I have presented an implementation of the common knowledge logic in Coq. There I have shown how this applies to prove mechanically popular (and less popular) puzzles as prolegomenon of other potential applications. In these experiments I have shown in particular that in the literature (mostly devoted to study *model theory* of common knowledge logic) some concepts of proof theory are not clearly brought out and statements made at the meta-level, i.e., in the meta-theory, are not sorted out from statements made at the level of the language, i.e., in the theory. In the deep embedding in a proof assistant (where the logic is fully implemented into the meta-language) the distinction between meta-theory and theory is made explicit, by construction. The proof assistant cannot accept ill-formed expressions and forces the user to specify the level of statements he makes, namely *inside* the theory or *outside* the theory. Thus the kind of implication or quantification or even statement, e.g., axiom or premise of a logical implication, has to be made precise. On the opposite, in the handwritten treatments of the puzzles, it is not clear whether a statement is made an axiom stated as such in the meta-theory or a proposition stated as the premise of a logical implication. This confusion is especially present in the literature on economic games [21, 8]. Using a quantification in the meta-theory vs a quantification in the theory can change dramatically the strength of a statement and its scope.

In this paper, my approach is this of a proof theorist with inclination to experiments. My goal is twofold. First I present a new axiomatization of common knowledge logic (axiom **FB** and rule **LFB**). Second I discuss a specific problem of common knowledge logic, namely the dilemma between internalizing or externalizing implication. Here one needs some explanation. In a proof theoretic approach there are two kinds of implications: an internal implication (the implication of the object theory) written here $? \Rightarrow ?$, and the external implication (the implication of the meta-theory) written $\frac{?}{?}\!\!\vdash$. Here $,\vdash$ means *" is a theorem"*. This discussion about the two views of the same problem in common knowledge logic will be made first

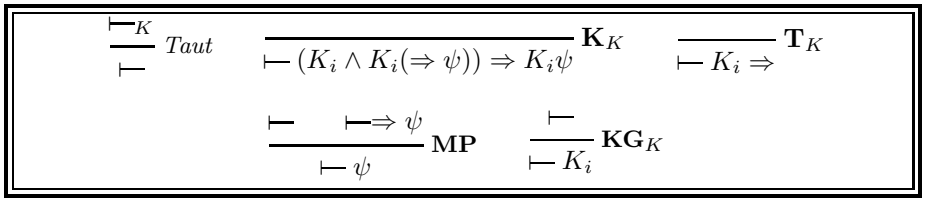

Figure 1: The basic rules of epistemic logic: the system $\mathbb{T}$

through examples and at this exploratory state no meta-theorem is proved. There are two approaches when solving a puzzle. In the first approach, a statement is made an axiom, say $\vdash$, this axiom leads to the proof of $\vdash \psi$, proving the meta implication $\frac{\vdash}{\vdash \psi}$. In the second approach, one proves $\vdash C_G() \Rightarrow \psi$, where C_G is the *common knowledge* modality. From experiments, I have drawn the following statements. These two approaches seem to be equivalent and show the interplay between the theory and the meta-theory. An interesting meta-theorem could be to prove that equivalence (see Section 5). I call *external vs internal* the equivalence of $\frac{\vdash}{\vdash \psi}$ with $\vdash C_G() \Rightarrow \psi$. In this paper all the discussion is based on experiments made in the proof assistant Coq and the paper can be seen as the description of those experiments. I discovered in [5] that the correspondence between $\frac{\vdash}{\vdash \psi}$ and $\vdash C_G() \Rightarrow C_G(\psi)$ is known, but it is not the one I am looking for. In what follows, the typewriter font is for code taken from the Coq implementation. Most of the development in Coq is available on the WEB at `http://perso.ens-lyon.fr/pierre.lescanne/COQ/epistemic_logic.v8` (see [13] or a presentation). The rest can be found in [19].

2 Presentation of common knowledge logic

Historical facts

The concept of common knowledge has been introduced by the philosopher Lewis [14] and since is used in several context namely distributed systems [12, 18], artifical intelligence [16] and game theory [1].

Epistemic logic

The basis of common knowledge logic is epistemic logic. In my experiments in Coq [4], epistemic logic is presented by a Hilbert-style system of rules and axioms. Since I use second order logic, I define only the (internal) implication and I derive the other connectors. There are only two rules namely **MP**, i.e., the *Modus Ponens* and **KG** also known as *Knowledge Generalization* and three axioms *Taut*, **K** and **T**. Actually *Taut* is an axiom scheme as it says that every classical tautology is a theorem in common knowledge logic. Such an approach requires a "deep embedding" (see annex A). The main reason is that modal logic cannot be easily implemented with natural deduction without changing its basic philosophy (see annex B). Epistemic logic is based on modal logic and in this paper only the system $\mathbb{T}$ (see Figure 1) is considered. Since there is much flexibility in the terminology, I decided to stick to the terminology of [5]. Epistemic logic introduces one modality for each agent: it expresses that that agent "knows" the proposition that follows the modality. More specifically, if is a proposition, $K_i()$ is the proposition modified by the modality K_i which means *"Agent i knows "*. In Figure 1, the statement $\vdash_K$ means that is a theorem in classical propositional logic (this time, K stands for the German adjective "klassisch" [9]). Knowing whether classical logic is relevant is a topics of research with René Vestergaard.

Common knowledge logic

Now let us suppose that we have a group G of agents. The knowledge of a fact can be shared by the group G, i. e., *"each agent in G knows "*. We write $E_G()$ and the meaning of E_G is easily axiomatized by the equivalence given in Figure 2 which can also be seen as the definition of E_G; it is called *shared knowledge*.

$$\frac{}{\vdash E_G() \Leftrightarrow \bigwedge_{i \in G} K_i}\ \mathbf{E}$$

Figure 2: Shared knowledge

In common knowledge logic, there is another modality, called *common knowledge* which is much stronger than shared knowledge. It is also associated with a group G of agents and is written C_G. Given , $C_G()$ is the least solution of the equation

$$x \quad \Leftrightarrow \quad \wedge E_G(x).$$

"Least" should be taken w.r.t. the order induced by $\Rightarrow$. A proposition ψ is *less than* a proposition ρ if $\rho \Rightarrow \psi$. As well known in the fixed point theory, the least solution of the above equation is also the least solution of the inequation:

$$x \quad \Rightarrow \quad \wedge E_G(x).$$

The axiomatization of Figure 3 characterizes $C_G()$ by two properties. Together with the system $\mathbb{T}$ and the definition of E_G it forms the system $\mathbb{CK}_G$. It asserts two things.

1. $C_G()$ is a solution of the inequation $x \Rightarrow \wedge E_G(x)$, axiom **FB**,
2. If ρ is another solution of the inequation, then ρ implies $C_G()$, which means that ρ is greater than $C_G()$). This is rule **LFB**.

One can prove that C_G satisfies axioms and rules of $\mathbb{T}$, where K_i is replaced by C_G even when $G = \emptyset$. Thus we prove

$$\frac{}{\vdash (C_G \wedge C_G(\Rightarrow \psi)) \Rightarrow C_G\psi}\ \mathbf{K}_C \qquad \frac{}{\vdash C_G \Rightarrow}\ \mathbf{T}_C \qquad \frac{\vdash}{\vdash C_G}\ \mathbf{KG}_C$$

$\mathbf{KG}_C$ stands for *Common Knowledge Generalization*. Notice that $\mathbf{T}_C$ and $\frac{\vdash}{\vdash}$ on one side and $\vdash C_G \Rightarrow C_G$ and $\mathbf{KG}_C$ on the other side form the two first instances of *external vs internal*. Actually one can prove more, namely that C_G satisfies axiom $\mathbf{4}_C$, namely $\vdash C_G() \Rightarrow C_G(C_G())$. It is a variant for common knowledge logic of the axiom $\vdash K_i() \Rightarrow K_i(K_i())$ of epistemic logic known as *Positive Introspection* or $\mathbf{4}_K$. The proof of $\mathbf{4}_C$ does not requires this of $\mathbf{4}_K$[1].

Notice that the presentation of common knowledge given in Figure 3 is new. It is more robust than this of Fagin et al. [5] which itself formalizes this of Aumann [1]. Our axiomatization works even for an empty set of agents and this is crucial, as starting with an empty set of agents is the key of a recursive definition of E_G and C_G;

Two presentations of common knowledge logic

This presentation should be compared with this given by Meyer and van der Hoek on page 46 of [17] (see Figure 4). The system $\mathbb{T} \cup \{A7, A8, A9, A10, R3\}$, together with the definition of E_G, is called $\mathbb{TEC}_G$. One notices that axioms $(A7)$ and $(A8)$ are just a splitting of axiom *Fixpoint*, i.e., one splits the conclusion $\wedge E_G(C_G())$. Axiom $(A9)$ is axiom $\mathbf{K}_C$ mentioned above and $(R3)$ is $\mathbf{KG}_C$ also mentioned above. As said, both $(A9)$ and $(R3)$ can be proved as theorems in $\mathbb{CK}_G$. $(A10)$ is more interesting and requires specific consideration. Figure 5 sketches a proof of $(A10)$ as a theorem in $\mathbb{CK}_G$. Therefore $\mathbb{CK}_G$ implies $\mathbb{TEC}_G$.

[1] This seems to show that **4**, which is a controverted axiom in general, should be stated more appropriately for the common knowledge of a group of agents rather than for the knowledge of an individual agent.

$$\frac{}{\vdash C_G() \Rightarrow \wedge E_G(C_G())}\ \mathbf{FB}$$

$$\frac{\vdash \rho \Rightarrow \wedge E_G(\rho)}{\vdash \rho \Rightarrow C_G()}\ \mathbf{LFB}$$

Figure 3: The rules for common knowledge

$(A7)$	$C_G()$	$\Rightarrow$	
$(A8)$	$C_G()$	$\Rightarrow$	$E_G(C_G())$
$(A9)$	$C_G() \wedge C_G(\Rightarrow \psi)$	$\Rightarrow$	$C_G(\psi)$
$(A10)$	$C_G(\Rightarrow E_G())$	$\Rightarrow$	$\Rightarrow C_G()$
$(R3)$	$\frac{}{C_G()}$		

Figure 4: Meyer and van der Hoek axioms $\mathbb{TEC}_G$

$\mathbb{TEC}_G$ ***implies*** $\mathbb{CK}_G$**.**

Indeed axiom **FB** is an obvious consequence of $\mathbb{TEC}_G$ and we prove that rule **LFB** is a consequence of $\mathbb{TEC}_G$ as follows.

$$\dfrac{\dfrac{\dfrac{\dfrac{\rho \Rightarrow \wedge E_G(\rho)}{\rho \Rightarrow E_G(\rho)}\,(R3)}{C_G(\rho \Rightarrow E_G(\rho))}\,(A10+\mathbf{MP})}{\rho \Rightarrow C_G(\rho)} \qquad \dfrac{\dfrac{\dfrac{\rho \Rightarrow \wedge E_G(\rho)}{\rho \Rightarrow}\,(R3)}{C_G(\rho \Rightarrow))}\,(A9+\mathbf{MP})}{C_G(\rho) \Rightarrow C_G()}}{\rho \Rightarrow C_G()}\,(Transitivity of \Rightarrow)$$

$(R10)$ ***implies*** $(A10)$**.**

In the above proof, we should notice that instead of axiom $(A10)$, we use rule

$$\frac{C_G(\Rightarrow E_G())}{\Rightarrow C_G()}\,(R10)$$

which is a direct consequence of $(A10)$ by **MP**. By analogy with $(A10)$, we call that rule $(R10)$. A closer look shows that we use the derived rule

$$\frac{\Rightarrow E_G()}{\Rightarrow C_G()}\,(R10')$$

which is the above rule combined with $(R3)$. See section *Discussion* below to understand why we are interested in that rule. Let us come back to $(R10)$ and let us call $\mathbb{TEC}'_G$ the system $\mathbb{T} \cup \{A7, A8, A9, R10, R3\}$. Since we have a proof of $\mathbb{CK}_G$ in $\mathbb{TEC}'_G$ and a proof of $\mathbb{TEC}_G$, in particular of $(A10)$, in $\mathbb{CK}_G$, we have an indirect proof of $\mathbb{TEC}_G$ in $\mathbb{TEC}'_G$ or, in short, of $(R10)$ implies $(A10)$. Here is a direct proof.

Let us state $A \equiv C_G(\Rightarrow E_G())$ in this proof. First, let us prove $A\wedge \Rightarrow C_G(A\wedge)$.

$$\dfrac{\dfrac{C_G(\Rightarrow E_G()) \Rightarrow E_G(C_G(\Rightarrow E_G()))\ \boxed{\text{A8}} \qquad \dfrac{C_G(\Rightarrow E_G()) \Rightarrow (\Rightarrow E_G())\ \boxed{\text{A7}}}{C_G(\Rightarrow E_G())\wedge \Rightarrow E_G()}}{C_G(\Rightarrow E_G())\wedge \ \Rightarrow\ E_G(C_G(\Rightarrow E_G())) \wedge E_G()}}{\dfrac{C_G(\Rightarrow E_G())\wedge \Rightarrow \qquad C_G(\Rightarrow E_G())\wedge \ \Rightarrow\ E_G(C_G(\Rightarrow E_G()))\wedge)}{C_G(\Rightarrow E_G()) \wedge \ \Rightarrow\ \wedge E_G(C_G(\Rightarrow E_G()))\wedge)}\ \textit{Transitivity of } \Rightarrow}\ \textbf{LFB}$$
$$\dfrac{C_G(\Rightarrow E_G())\wedge \Rightarrow C_G()}{C_G(\Rightarrow E_G()) \Rightarrow\Rightarrow C_G()}$$

Figure 5: A proof of Meyer and van der Hoek's axiom ($A10$)

$$\dfrac{\dfrac{C_G(\Rightarrow E_G()) \Rightarrow E_G(C_G(\Rightarrow E_G()))\ \boxed{\text{(A8)}} \qquad \dfrac{C_G(\Rightarrow E_G()) \Rightarrow (\Rightarrow E_G())\ \boxed{\text{(A7)}}}{C_G(\Rightarrow E_G())\wedge \Rightarrow (\Rightarrow E_G())\wedge}}{A\wedge \Rightarrow E_G(A)} \qquad \dfrac{(\Rightarrow E_G())\wedge \ \Rightarrow\ E_G()}{C_G(\Rightarrow E_G())\wedge \Rightarrow E_G()}}{\dfrac{A\wedge \ \Rightarrow\ E_G(A\wedge)}{A\wedge \Rightarrow C_G(A\wedge)}\,(R10)}$$

The rest is easy. First, we notice that we have $C_G(A\wedge) \Rightarrow C_G()$.

$$\dfrac{\dfrac{A\wedge \Rightarrow}{C_G(A\wedge \Rightarrow)}\,(R3)}{C_G(A\wedge) \Rightarrow C_G()}\,(A9) + \mathbf{MP}$$

By transitivity of $\Rightarrow$, we get $A\wedge \Rightarrow C_G()$. But clearly $A\wedge \Rightarrow C_G()$ is equivalent to $A \Rightarrow\Rightarrow C_G()$ which is $C_G(\Rightarrow E_G()) \Rightarrow\Rightarrow C_G()$, e.g., $(A10)$.

Discussion

The equivalence between $(A10)$ and $(R10')$ is a third instance of *external vs internal*. Indeed, we have shown that a proposition of the form $\vdash C_G(\rho) \Rightarrow \psi$ is equivalent to a rule of the form $\frac{\vdash \rho}{\vdash \psi}$.

3 The three wise men

The first example we address is the well-known example of the three wise men. See [13] for a more detailed presentation. It is stated usually as follows ([5], Exercise 1.3): *"There are three wise men. It is common knowledge that there are three red hats and two white hats. The king puts a hat on the head of each of the three wise men and asks them (sequentially) if they know the color of the hat on their head. The first wise man says that he does not know; the second wise man says that he does not know; then the third man says that he knows"*. Let us call the three wise persons Alice, Bob and Carol. Let us write `white Alice` for *"Alice wears a white hat"* and `red Alice` for *"Alice wears a red hat"*. The puzzle is based on a function which says whether an agent knows the color of her (his) hat:

```
Definition Kh := fun i => (K i (white i)) V (K i (red i)).
```

Clearly one has to prove that `Kh Carol` holds under some assumptions. To make clear theses assumptions, we define in addition a few propositions namely

```
Definition One_hat := \-/(fun i:nat => white i | red i).
```

which says that every agent wears a red hat or a white hat. If P is a predicate, \-/P is the logical quantification, i.e., the quantification in the theory not this in the meta-theory.

```
Definition Two_white_hats := white Bob & white Carol ==> red Alice.
```

which says that there are two white hats. Notice that this is stated in a weak form, indeed it is only when Bob and Carol wear white hats that one can deduce that Alice wears a red hat. Moreover there are three concepts which say that each agent sees the hat of the other agents and therefore knows the color of the hat.

```
Definition K_Alice_white_Bob :=  white Bob ==> K Alice (white Bob).
Definition K_Alice_white_Carol :=  white Carol ==> K Alice (white Carol).
Definition K_Bob_white_Carol :=  white Carol ==> K Bob (white Carol).
```

A first result

In a first attempt [13], the five above propositions were stated as axioms and I was able to prove:

```
|- K Carol (K Bob (¬ Kh Alice) & ¬ Kh Bob)
            ==> K Carol (red Carol).
```

In Coq this would give a statement like

```
|- One_hat &
   K_Alice_white_Bob &
   K_Alice_white_Carol &
   K_Bob_white_Carol &
   Two_white_hats  ->
|- K Carol (K Bob (¬ Kh Alice) & ¬ Kh Bob)
           ==> K Carol (red Carol).
```

where -> is the meta-implication, i.e., this of Coq and as usual |− says that proposition is a theorem.

A second result

In the second attempt one proves:

```
|- K Carol (K Bob (One_hat &
                   K_Bob_white_Carol &
                   K_Alice_white_Bob &
                   K_Alice_white_Carol &
                   (K Alice Two_white_hats) &
                   ¬ Kh Alice) &
             ¬ Kh Bob)
  ==> Kh Carol.
```

This tells exactly the amount of knowledge which Carol requires to deduce that she knows the color of her hat, actually red. Let us call Alice_Bob_Carol the group made of Alice, Bob and Carol. From the above statement, one derives the corollary:

```
|- C Alice_Bob_Carol (Two_white_hats &
                       One_hat &
                       K_Bob_white_Carol &
                       K_Alice_white_Bob &
                       K_Alice_white_Carol)
  ==> K Carol (K Bob (¬ Kh Alice) & ¬ Kh Bob) ==> Kh Carol.
```

which is weaker. But if we state

$\varphi \equiv$
```
Two_white_hats &
One_hat &
K_Bob_white_Carol &
K_Alice_white_Bob &
K_Alice_white_Carol
```

and

$\psi \equiv$ `K Carol (K Bob (¬ Kh Alice) & ¬ Kh Bob) ==> Kh Carol`

we notice that we have exhibited a fourth instance of *external vs internal* since $\vdash C_G() \Rightarrow \psi$ and $\frac{\vdash}{\vdash \psi}$ are equivalent.

4 The muddy children

This problem had many variants [15, 7, 6, 8]. It is a typical example of how a community of agents acquires knowledge. In its politically correct version [5, 17], a group of children have mud on their head after playing during a birthday party. The kids do not know they have mud on their head. The father of the kid who organized the party asked the children to come around him in a circle for the kids to see each other and he tells them that there is at least one child who has mud on his face so that they clearly all hear him. Then Father asks the kids who have mud to step forward. He repeats this last sentence until all the kids step forward.

Philosophers have been puzzled by the fact that the first sentence of Father namely *"There is at least one child with mud on his face"* is absolutely necessary. This fact is known by the children, but by doing so, Father makes it a common knowledge. In [13], we have identified that the key lemma is

```
Lemma Progress :
    forall n p : nat,
    |- C ([:n+1:]) (At_least (n+1) p) &
       E ([:n+1:]) (¬ Exactly (n+1) p)
       ==> C ([:n+1:]) (At_least (n+1) (p+1)).
```

In other words, if the fact that there is at least p muddy children is a common knowledge and all the children know that there is not exactly p muddy children, then the fact that there is at least $p+1$ muddy children is a common knowledge. Together with the first statement of Father:

```
Axiom First_Father_Statement :
    |- C ([:nb_children:]) (At_least n 1).
```

we are able to prove after n steps `C ([:n:]) (At_least n n)` which means that *the fact that there is at least n muddy children is common knowledge*. This is the final result. Common knowledge is important here because one can "progress" in common knowledge and not in shared knowledge. Thus the first statement that provides a first common knowledge allows initialization. The proof of Progress relies on a statement

```
Knowledge_Diffusion :
 forall n p i : nat,
    |- E ([:n:]) (At_least n p) ==>
       E ([:n:]) (¬ Exactly n p) ==>
       K i (E ([:n:]) (¬ Exactly n p)).
```

This statement is here to translate what children see after Father has asked the muddy ones to step forward and none did. They all know that there is at least p muddy children and they all know that there is not exactly p muddy children otherwise those with muddy face would have stepped forward, but now each one knows that all the others know that there is not exactly p muddy children.

Knowledge_Diffusion as an axiom

In a first experiment, we made `Knowledge_Diffusion` an axiom and we were able to prove `Progress` in its above form.

Knowledge_Diffusion as a common knowledge

In the second experiment, we consider that proposition `Knowledge_Diffusion` should not be made an axiom, i.e., an immutable principle, but it should be made just a rule of a game upon everyone agrees. Therefore the rules of the game are common knowledge that everyone accepts; agreeing on these rules makes everyone to act and reason according to them, i.e., "rationally". In this version *Progress* becomes:

```
 Lemma Progress :
forall n p : nat,
|- C ([:n+1:])(Knowledge_Diffusion) ==>
   (C ([:n+1:]) (At_least (n+1) p) &
    E ([:n+1:]) (¬ Exactly (n+1) p))
   ==> C ([:n+1:]) (At_least (n+1) (p+1)).
```

Discussion

Again we show that we can change an statement of the form $\frac{\vdash}{\models_\psi}$ into a statement of the form $\vdash C_G() \Rightarrow \psi$. Here

```
φ ≡ C ([:n+1:]) (At_least (n+1) p) &
    E ([:n+1:]) (¬ Exactly (n+1) p))
```

and

```
ψ ≡ C ([:n+1:]) (At_least (n+1) (p+1)).
```

This is a fifth instance of *external vs internal*.

5 The equivalence between internal and external implication

Fagin et al [5] in exercise 3.29 notice, with no reference, that $\frac{\vdash}{\models_\psi}$ and $\vdash C_G() \Rightarrow C_G(\psi)$ are equivalent. One notice by $\mathbf{T}_C$, i.e., $\vdash C_G(\rho) \Rightarrow \rho$, that this statement is stronger than *external vs internal*, which states the equivalence between $\frac{\vdash}{\models_\psi}$ and $\vdash C_G() \Rightarrow \psi$. The proof of that result cannot be readily implemented in Coq in our current implementation of common knowledge logic since this requires a deeper embedding of the theory. In short, in order to mechanize that proof, one needs not only internalize the object implication, which we called internal implication, but also what we called the external implication, since a meta-proof of the equivalence requires an induction on the proof of $\frac{\vdash}{\models_\psi}$. In a first step, one can prove in Coq that all the rules of common knowledge logic, namely **MP**, **KG** and **LFB** have their equivalent in the form $\vdash C_G() \Rightarrow C_G(\psi)$, namely:

$$\vdash C_G((\Rightarrow \psi)\wedge) \Rightarrow C_G(\psi) \qquad \vdash C_G() \Rightarrow C_G(K_i())$$

$$\vdash C_G(\rho \Rightarrow \wedge E_G(\rho)) \Rightarrow C_G(\rho \Rightarrow C_G())$$

The first one is a variant, by the means of $\vdash C_G(\chi \wedge \rho) \Leftrightarrow C_G(\chi) \wedge C_G(\rho)$, of $\mathbf{K}_C$ or $(A9)$. The second one is a basic result of common knowledge logic. The third theorem has no equivalent in the literature and has been proved in Coq for that purpose. Then we get the following interesting result:

$$\vdash C_G(\varphi) \Rightarrow C_G(\psi) \longrightarrow \vdash C_G(\varphi) \Rightarrow \psi \longrightarrow \frac{\vdash \varphi}{\vdash \psi}$$

The back arrow is proved by induction of the length of the deduction $\vdash$ -> $\vdash \psi$. Therefore, one notices three levels of implications: the implication $\Rightarrow$ in the theory, the implication $\overset{\vdash?}{\Rightarrow}$ in the metatheory and the implication $\longrightarrow$ in the meta-metatheory. From the above diagram one gets

$$\vdash C_G(\varphi) \Rightarrow \psi \quad \longrightarrow \vdash C_G(\varphi) \Rightarrow C_G(\psi) \ .$$

Actually we have

$$\frac{\vdash C_G(\varphi) \Rightarrow \psi}{\vdash C_G(\varphi) \Rightarrow C_G(\psi)}$$

as follows

$$\dfrac{\dfrac{\vdash C_G(\varphi) \Rightarrow \psi \qquad \vdash C_G() \Rightarrow E_G(C_G())}{\vdash C_G() \Rightarrow \psi \wedge E_G(C_G())}\,\mathbf{LFB}}{\vdash C_G(\varphi) \Rightarrow C_G(\psi)}$$

since $\vdash C_G() \Rightarrow E_G(C_G())$ is a theorem of common knowledge logic.

6 Conclusion

On another hand, it is worth to mention the study on combining common knowledge logic and dynamic logic we have done with Jérôme Puisségur [20]. The dynamic logic is used to describe changes in the world, but those changes are *purely epistemic* (an idea we borrow from Baltag, Moss and Solecki [3, 2]). This means that they affect only knowledge of the agents and nothing else. The muddy children puzzle has been axiomatized in this framework and a proof of its results has been fully mechanized in Coq. We can draw already two lessons form those experiences. First when merging two modal logics it seems that internalizing common knowledge is more appropriate. In other words, an approach like $\vdash C_G() \Rightarrow \psi$ should be preferred to setting the axiom $\vdash$ to prove $\vdash \psi$, as one does not know which metatheory a specific statement belongs to: dynamic logic or common knowledge logic? Second a formalization of predicate logic, allows expressing easily arbitrary depth of shared logic according to the number of agents. More precisely, common knowledge is not a priori necessary in the muddy children example and just a specific number of imbricated shared knowledge modalities corresponding to the number of children. This fact was already noticed by authors [8].

Acknowledgment

I would like to thank Bertrand Prémaillon who made part of the experiments in Coq and René Vestergaard for stimulating discussions.

References

[1] R. J. Aumann. Backward induction and common knowledge of rationality. *Games and Economic Behavior*, 8:6–19, 1995.

[2] A. Baltag. A logic of epistemic actions. In W. van der Hoek, J.-J. Meyer, and C. Witteveen, editors, *Proceedings of the ESSLLI 1999 workshop on Foundations and Applications of Collective Agent-Based Systems*, Utrecht University, 1999.

[3] A. Baltag, L. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicion. In *Proc. of TARK*, pages 43–56. Morgan Kaufmann Publishers, 1998.

[4] Y. Bertot and P. Castéran. *Interactive Theorem Proving and Program Development Coq'Art: The Calculus of Inductive Constructions*. Springer-Verlag, 2004.

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

[6] G. Gamow and M. Stern. *Puzzle math*, chapter *Forthy unfaithful wives*, pages 20–23. The Viking Press, New York, 1958.

[7] M. Gardner. *Puzzles from other worlds*. Vintage, 1984.

[8] J. Geanakoplos. *Handbook of Game Theory, R. Aumann and S. Hart ed.*, volume 2, chapter *Common knowledge*, pages 1437–1496. Elsevier, Amsterdam, 1994.

[9] G. Gentzen. Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1935.

[10] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.

[11] Y. Lafont. From proof nets to interaction nets. In J.-Y. Girard, Y. Lafont, and L. Regnier, editors, *Advances in Linear Logic*, pages 225–247. Cambridge University Press, 1995.

[12] D. Lehmann. Knowledge, common knowledge and related puzzles (extended summary). In *PODC '84: Proceedings of the third annual ACM symposium on Principles of distributed computing*, pages 62–67, New York, NY, USA, 1984. ACM Press.

[13] P. Lescanne. Mechanizing epistemic logic with Coq. Research Report RR2004-27, LIP-ENS de Lyon, 2004. An updated version submiited for publication is avalaible at `http://perso.ens-lyon.fr/pierre.lescanne/PUBLICATIONS/common_knowledge`.

[14] D. Lewis. *Convention: A philosophical study*. Harvard University Press, Cambridge, MA, 1969.

[15] J. E. Littlewood. *Littlewood's miscellany*. Cambridge University Press, Cambridge, 1986.

[16] J. McCarthy, M. Sato, T. Hayashi, and S. Igarashi. On the model theory of knowledge. Technical Report AIM-312, Stanford University, 1977.

[17] J-J Ch. Meyer and W van der Hoek. *Epistemic Logic for Computer Science and Artificial Intelligence*, volume 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1995.

[18] Y. O. Moses, D. Dolev, and J. Y. Halpern. Cheating husbands and other stories: a case study in knowledge, action, and communication. *Distributed computing*, 1(3):167–176, 1986.

[19] B. Prémaillon. Logique épistémique, modélisation dans un assistant de preuve. Master's thesis, Master Ingénierie Mathématique, Université Claude Bernard, Lyon, 2005.

[20] J. Puisségur. Eléments de construction d'une logique épistémique et dynamique. Rapport de stage de licence de l'École normale supérieure de Lyon, 2005.

[21] D. Samet. Hypothetical knowledge and games with perfect information. *Games and Economic Behavior*, 17:230–251, 1996.

A Deep embedding

A logic $\mathcal{L}$, the object logic or the object theory, is said to be deeply embedded in another logic $\mathcal{M}$, the meta-theory, or in a proof assistant if one considers the logic $\mathcal{M}$ to be this of the proof assistant, if all the constituents of the logic $\mathcal{L}$ are made objects of the logic $\mathcal{M}$ and all the connectors and the rules of $\mathcal{L}$ are defined inside the logic $\mathcal{M}$. This is opposed to shallow embedding where $\mathcal{L}$ and $\mathcal{M}$ may share connectors and rules. A shallow embedding is usually more concise, but in a deep embedding a clear distinction is made between the connectors of the object theory and those of the meta-theory. In a deep embedding the connector

and the corresponding meta-connector can be somewhat connected, but they cannot match completely. For instance, it could happen that the meta-disjunctions of two propositions meta-implies the proposition made as the conjunction of the two propositions and not vice-versa, in a sense made precise in formalizing the object theory.

Moreover not all the logics can be shallowly embedded. This is the case for common knowledge logic which cannot be formalized easily in a natural deduction framework (see next section).

B Why an Hilbert approach?

The reason why one cannot use a natural deduction of a sequent calculus approach is essentially due to the rule **KG**. If one accepts such a rule in natural deduction, one gets

$$\frac{\Gamma \vdash}{K_i(\Gamma) \vdash K_i()}$$

This requires to extend the operator K_i to contexts like Γ. If instead of K_i one uses a modality $\Box$, one says that $\Box(\Gamma)$ is a *"boxed context"*. Actually *linear logic* [10] is perhaps the archetypical modal logic and the equivalent of K_i is the modality *of course* written "!". The equivalent of **KG** is a rule called also *of course*. Without that rule the proof net presentation is somewhat simple [11]. Its introduction requires a machinery of boxes which increases its complexity.