





He **believes** he is Napoleon,
but **it is well known**
that I am Napoleon.



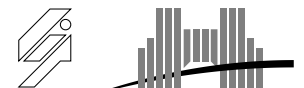
Epistemic Logic in Higher Order Logic
An experiment with COQ

Pierre Lescanne, LIP, ENS de Lyon

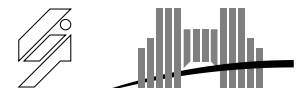
”Reports that say something hasn’t happened are always interesting to me, because as we know, there are known knowns; there are things we know we know,”

”We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don’t know we don’t know.”

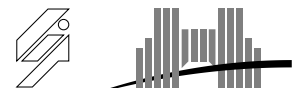
Defense Secretary Donald Rumsfeld,
at a news briefing in February 2002



Examples related to computers

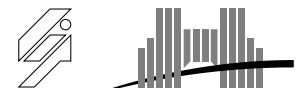


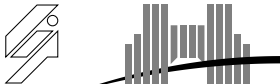
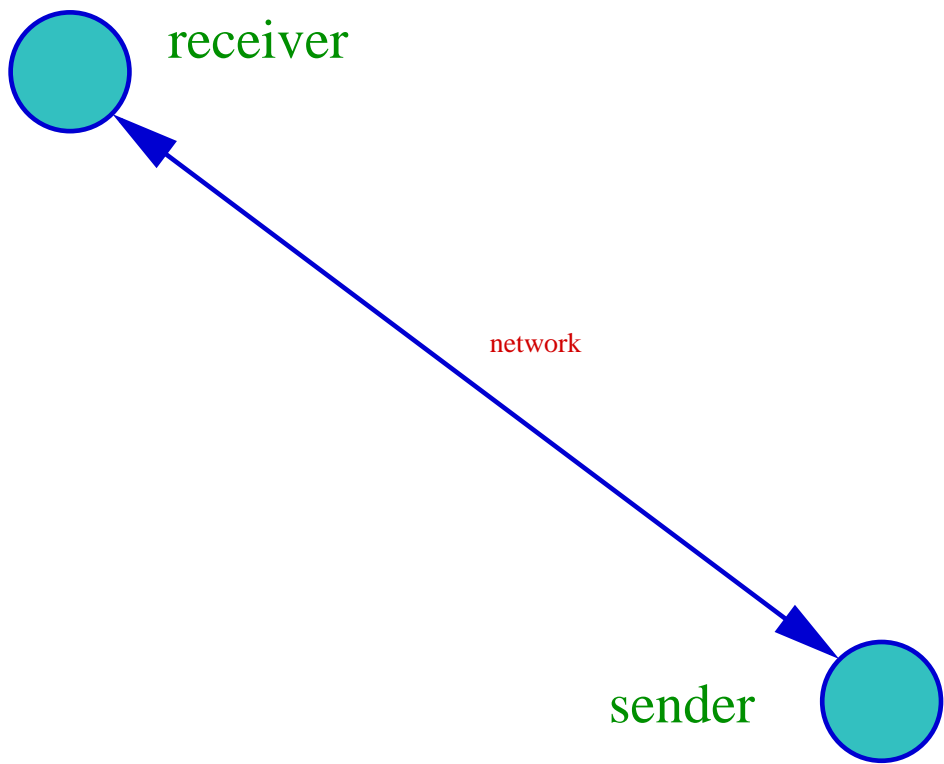
A sender receiver protocol



A sender receiver protocol

Network transmits messages between a sender and a receiver:



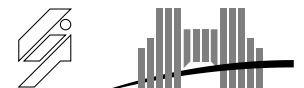


A sender receiver protocol

Network transmits messages between a sender and a receiver:

- network **can duplicate** messages,
- network **can loose** messages,
- however, network cannot **loose** a message **forever**.

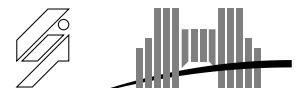
This is **Internet TCP**.



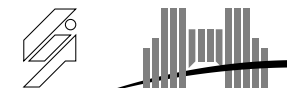
A sender receiver protocol (*suite*)

As long as the sender **does not know** whether the receiver has received a given message m_i , it resends it.




The receiver acknowledges reception of a message by sending an **acknowledgment** message ack_i as long as it **does not know** whether the sender has received this acknowledgment.



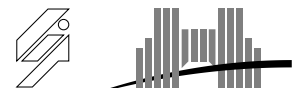
The coordinated attack

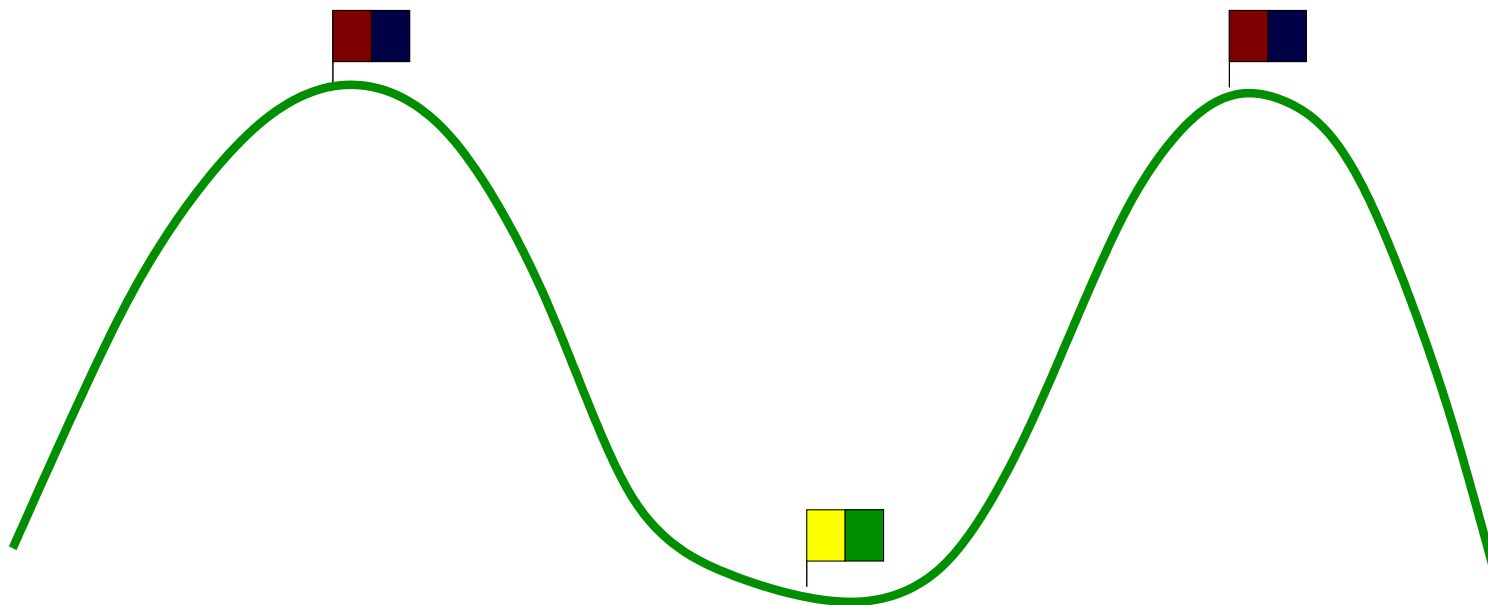


The coordinated attack

- Two generals  and their armies on two hills,
- They must attack **together** the enemy , i.e., **at the same hour**.
- Each general must be sure that the other will attack at the same time.
- **They communicate through messengers** 
 - who take half an hour to go from one camp to the other,
 - who can be caught, be killed or get lost.

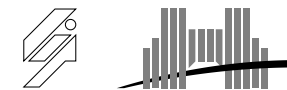
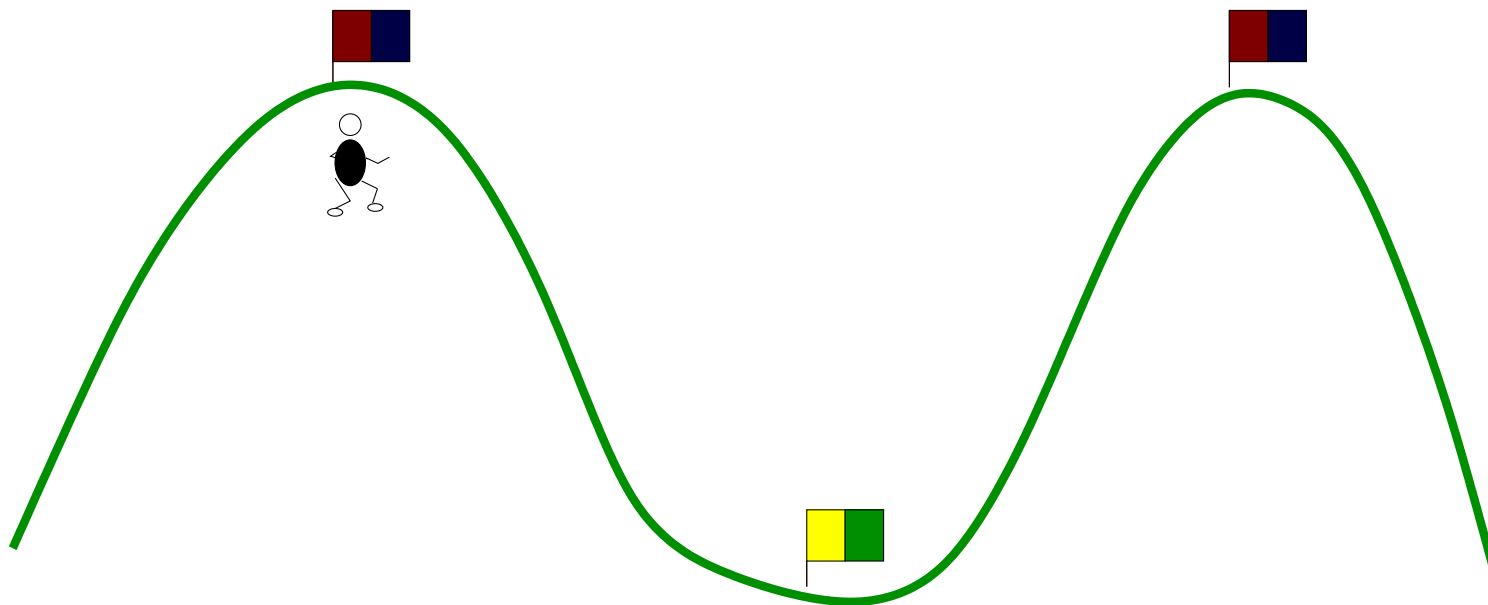
How do the generals coordinate their attack?

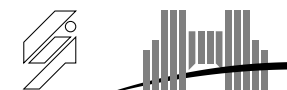
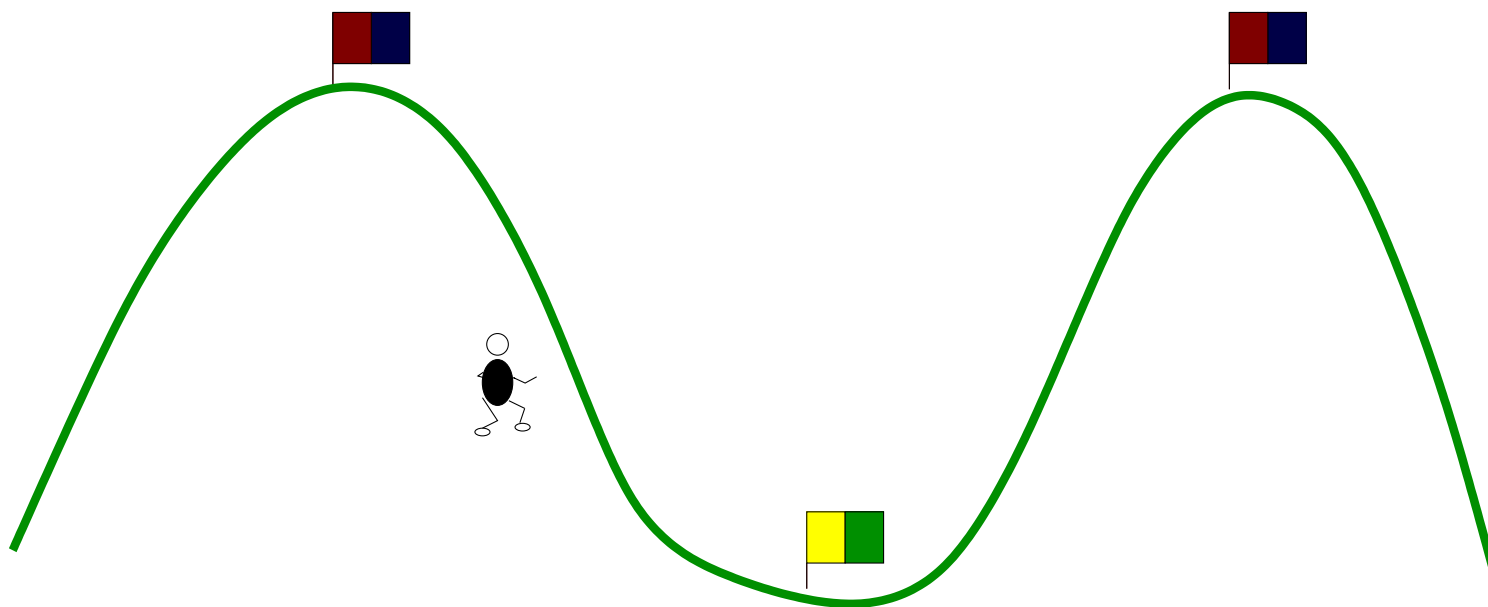


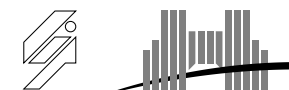
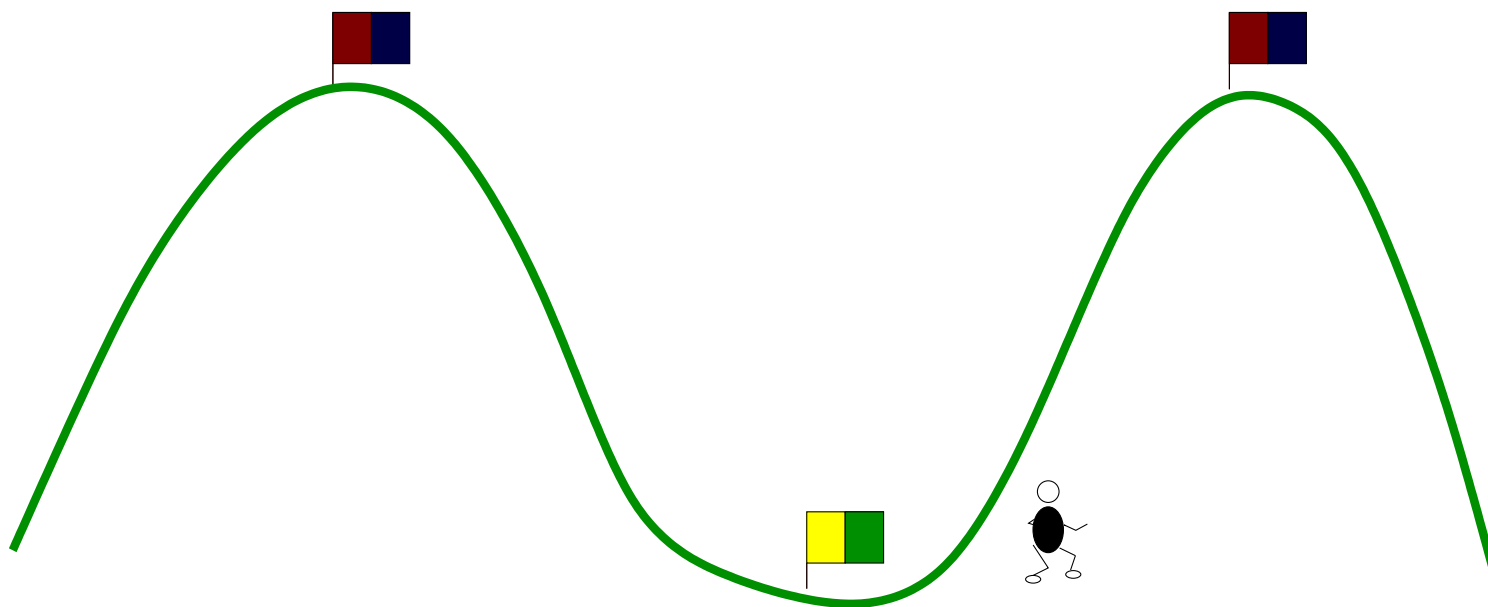


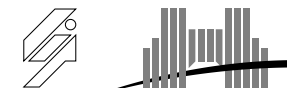
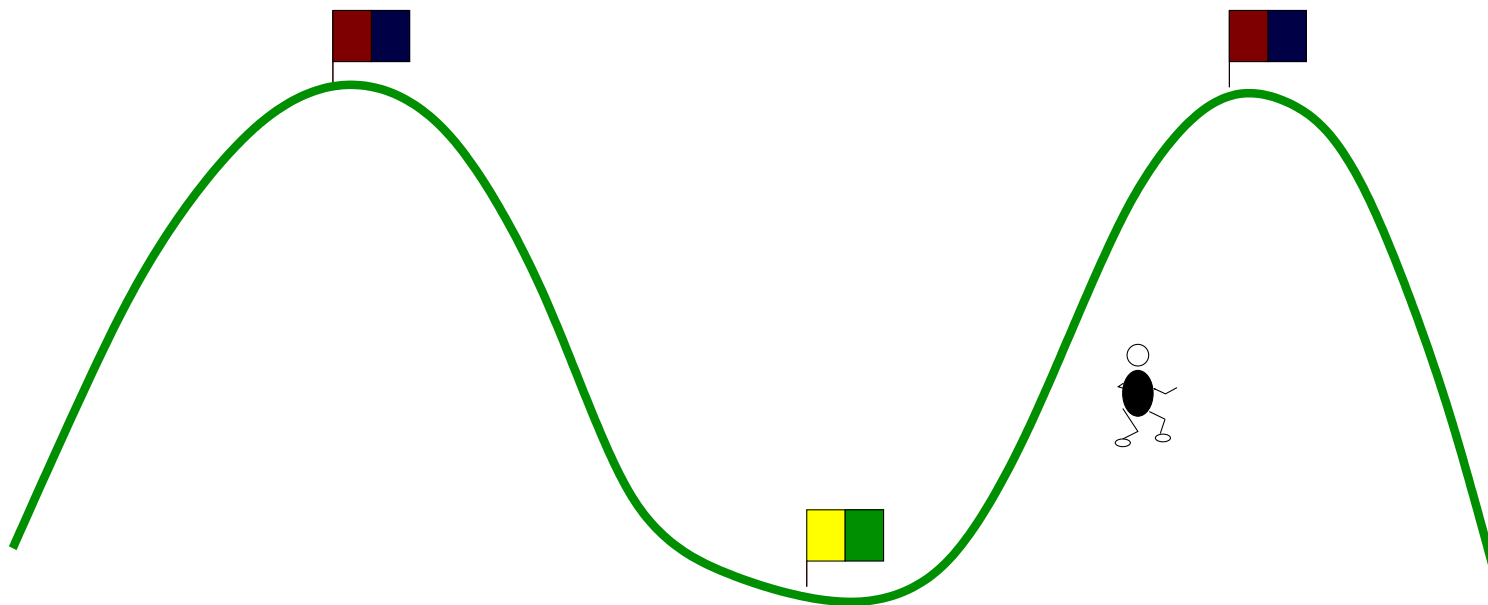
March 10, 2004

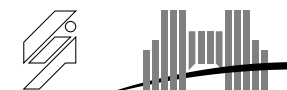
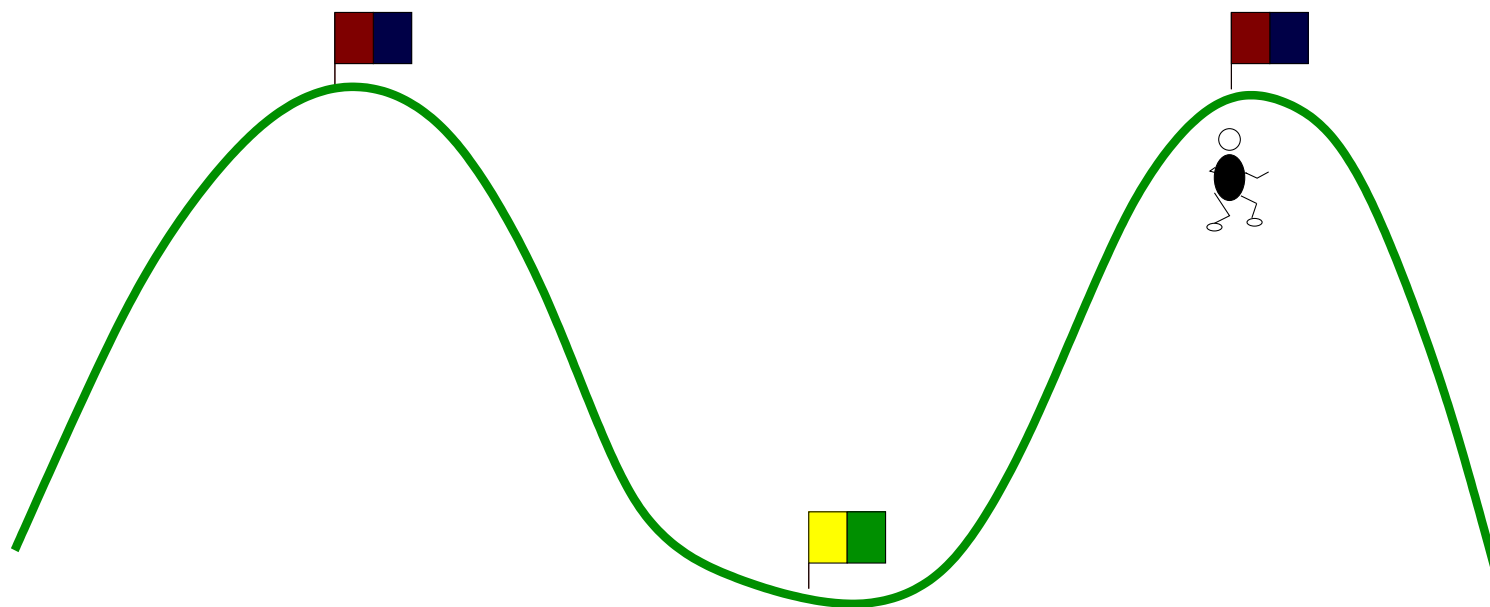




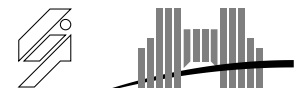


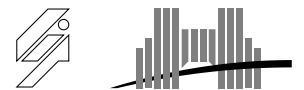
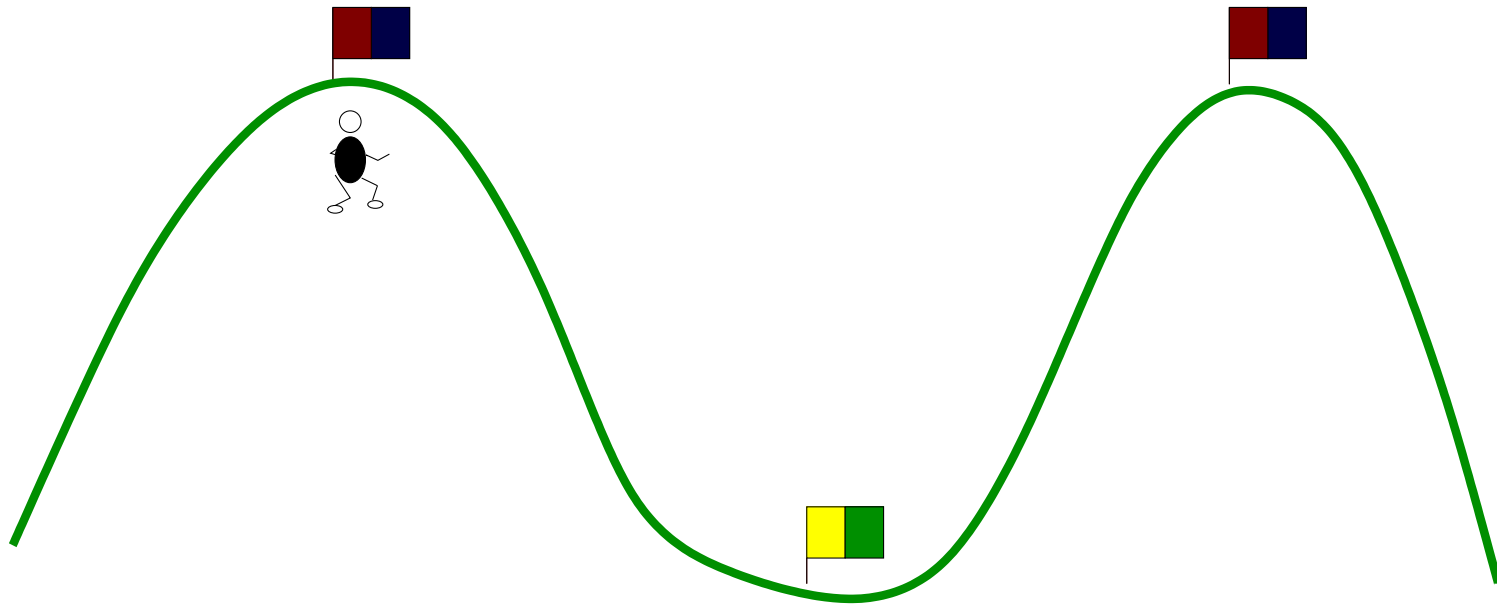


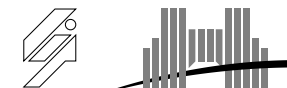
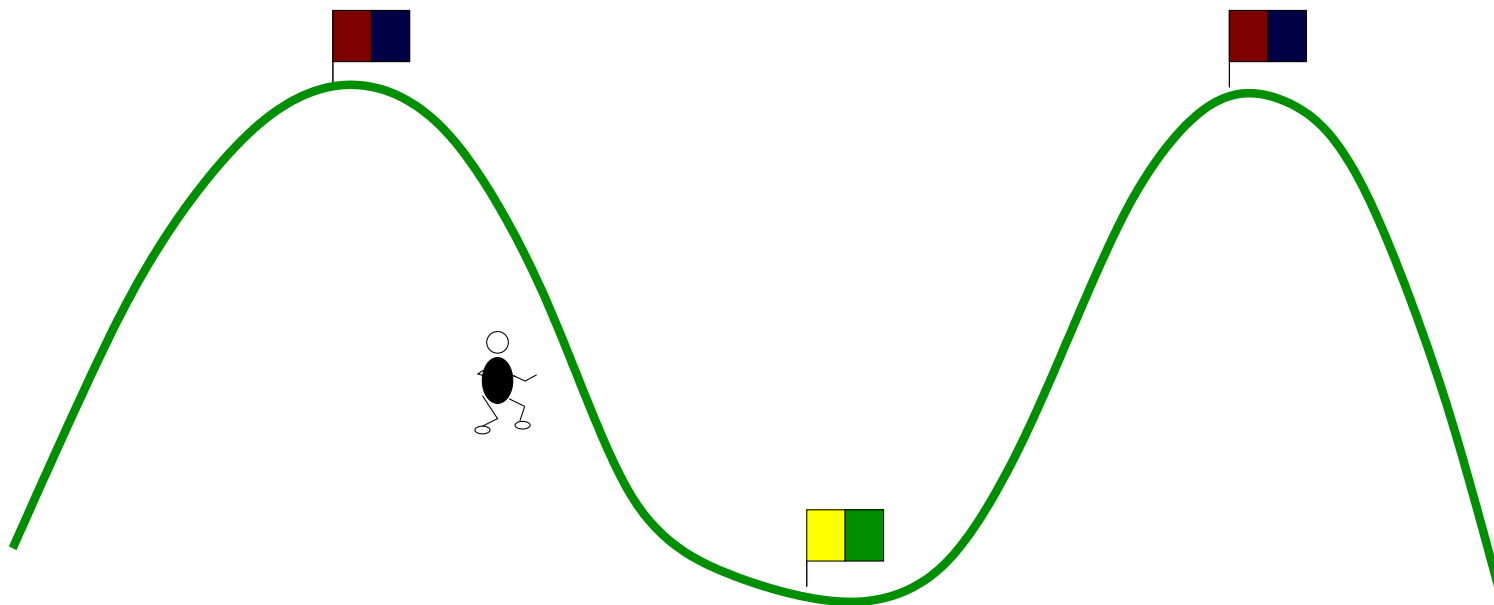


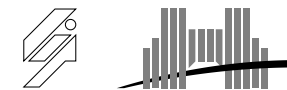
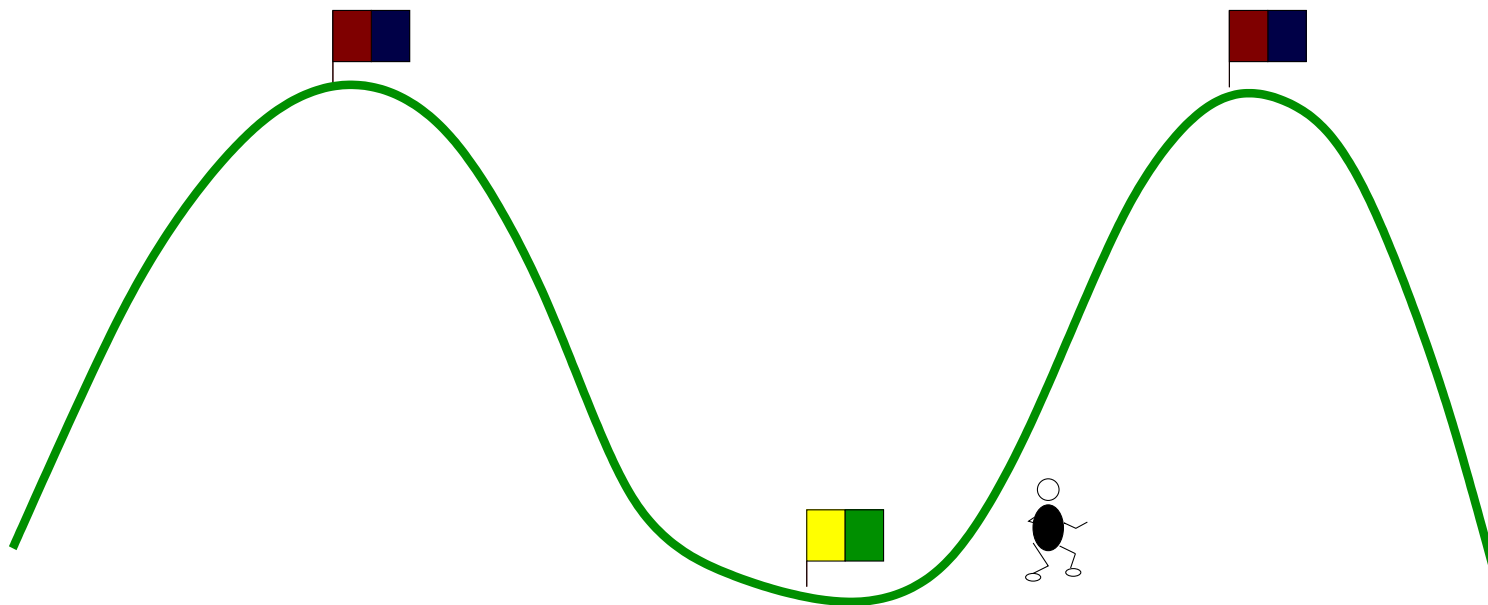


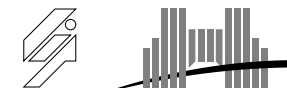
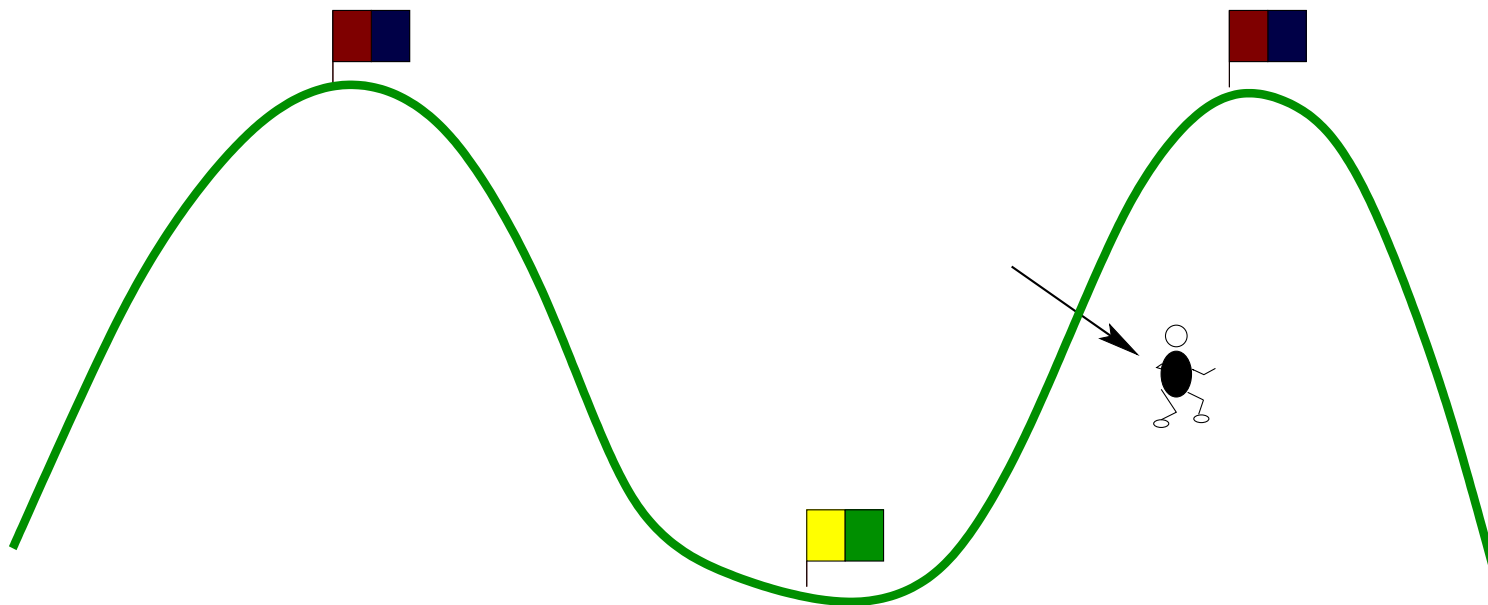
***But, the messenger can
be caught or killed !***

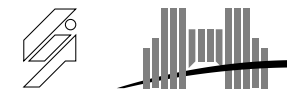
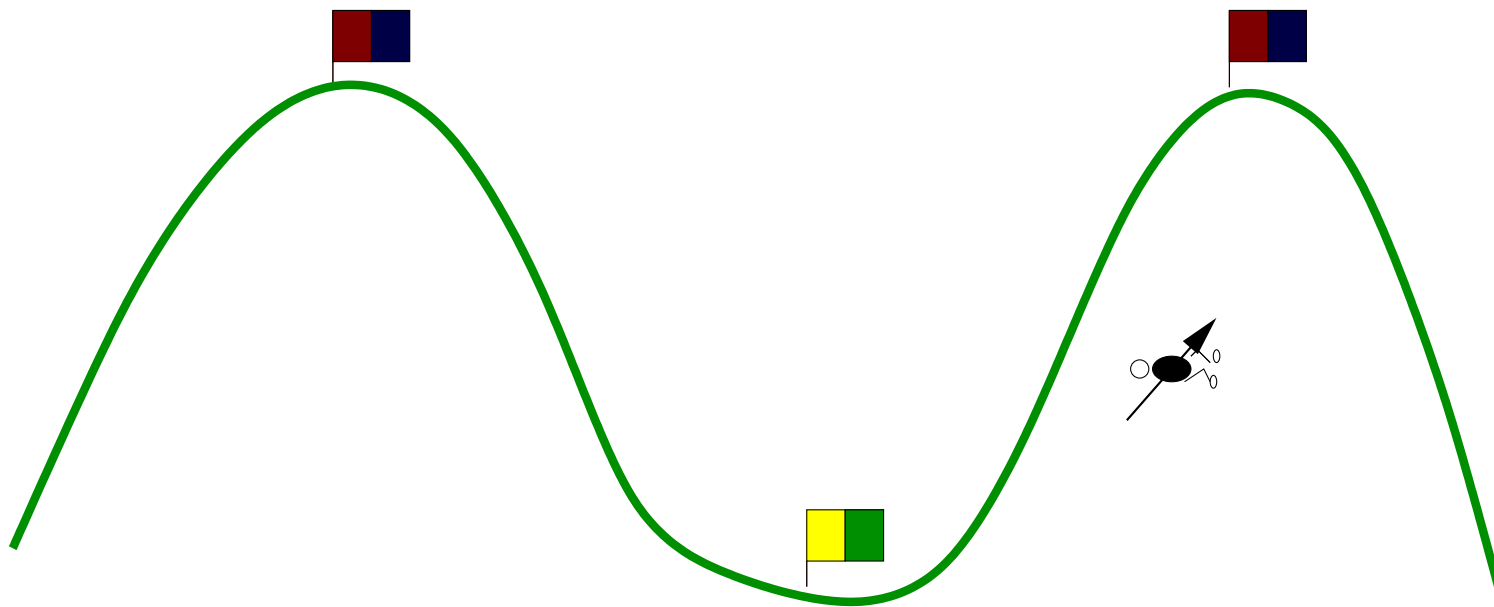




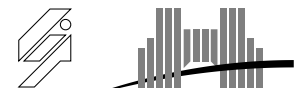


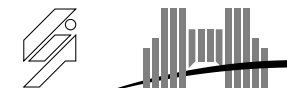
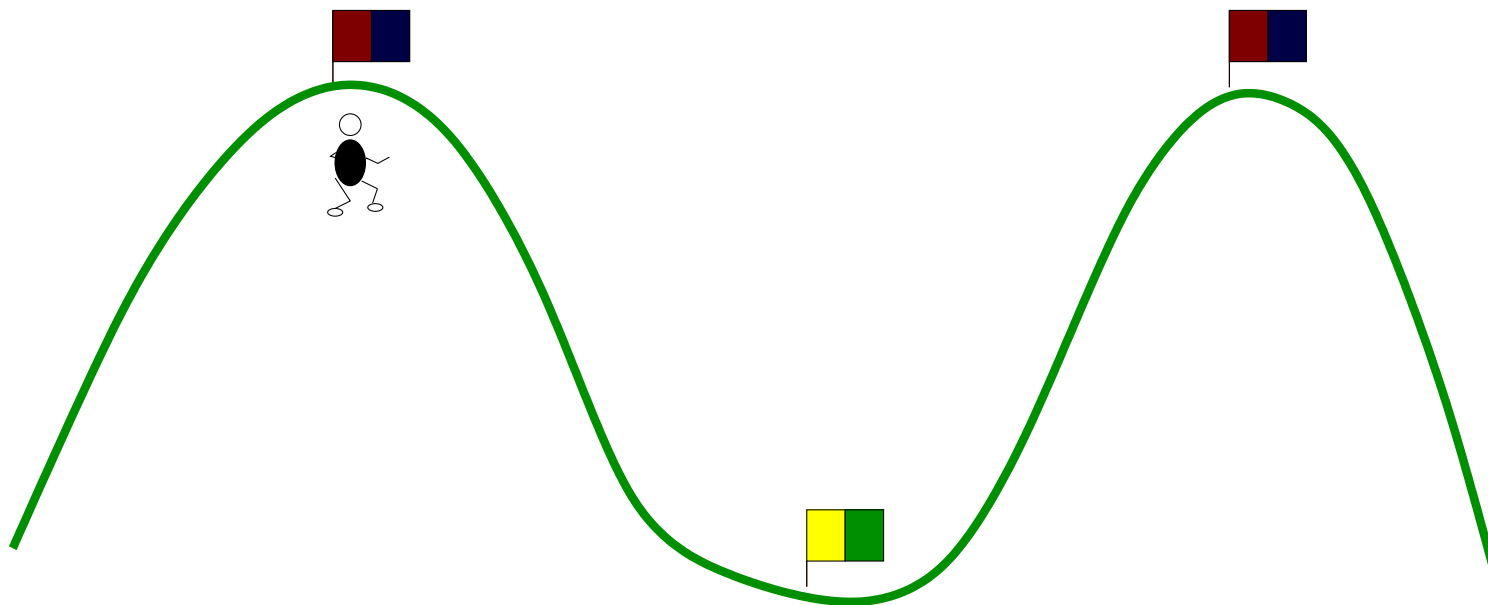


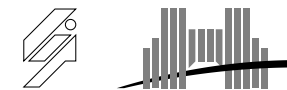
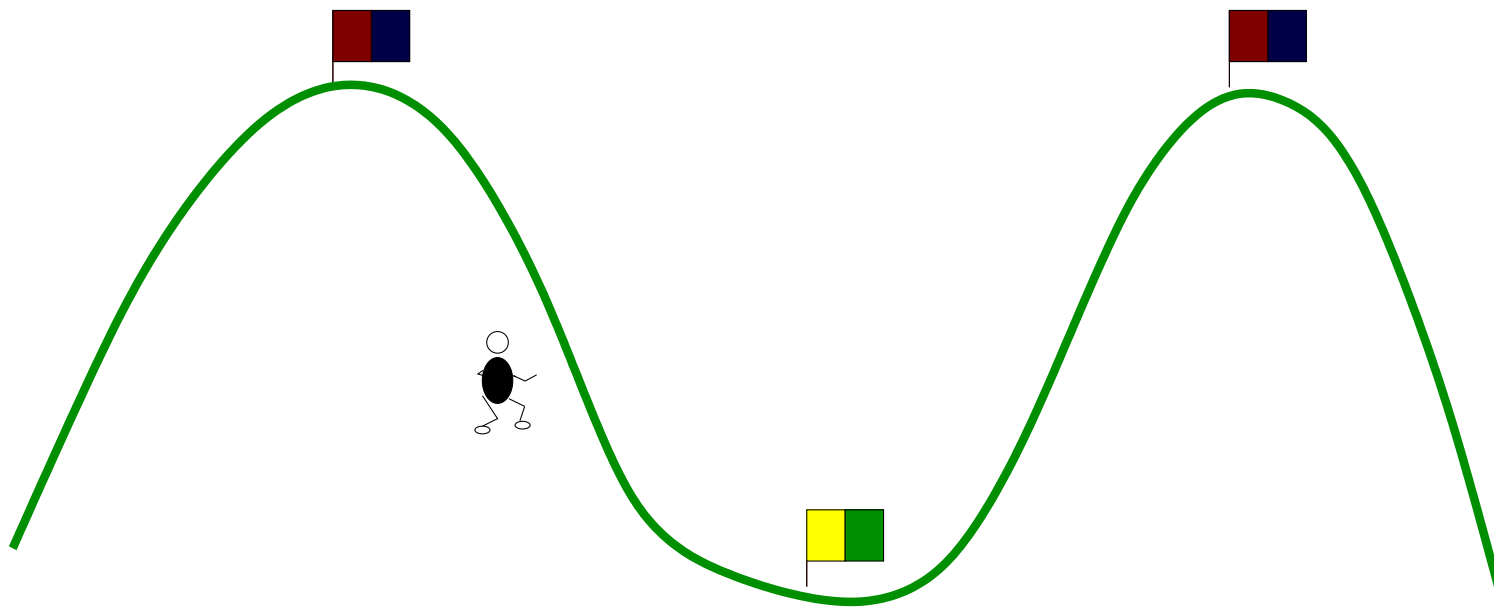


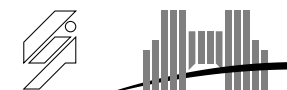
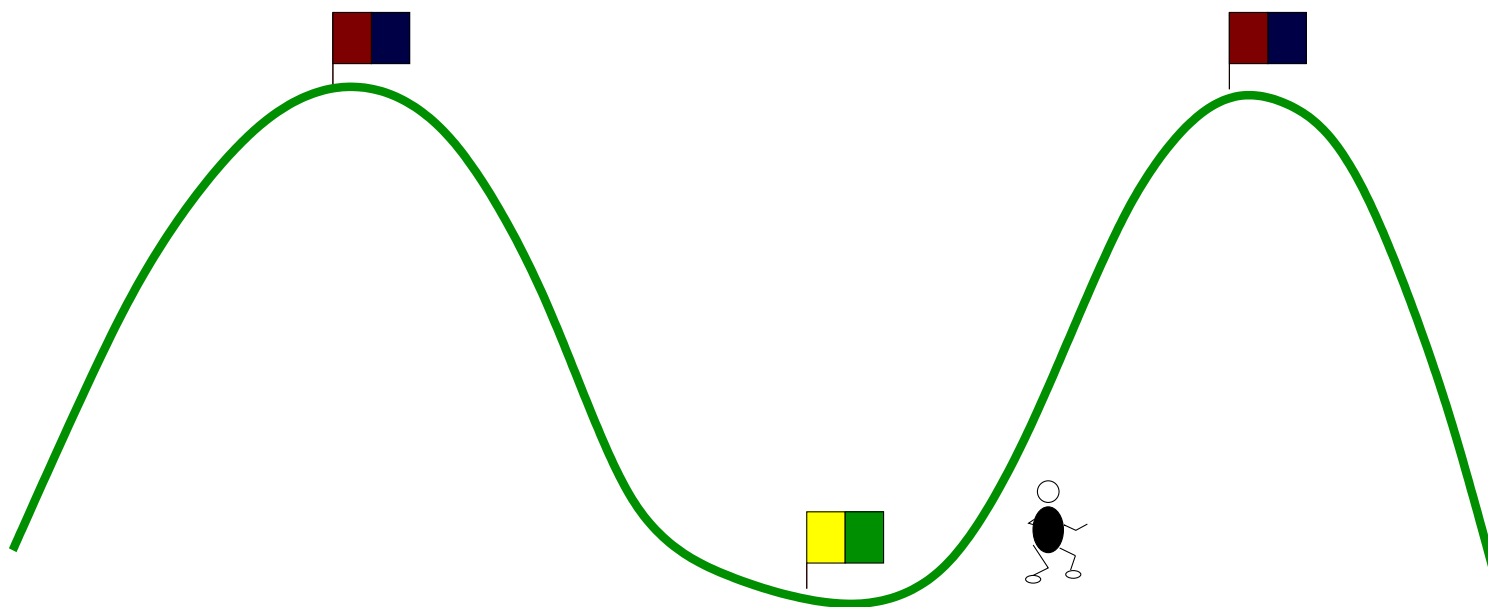


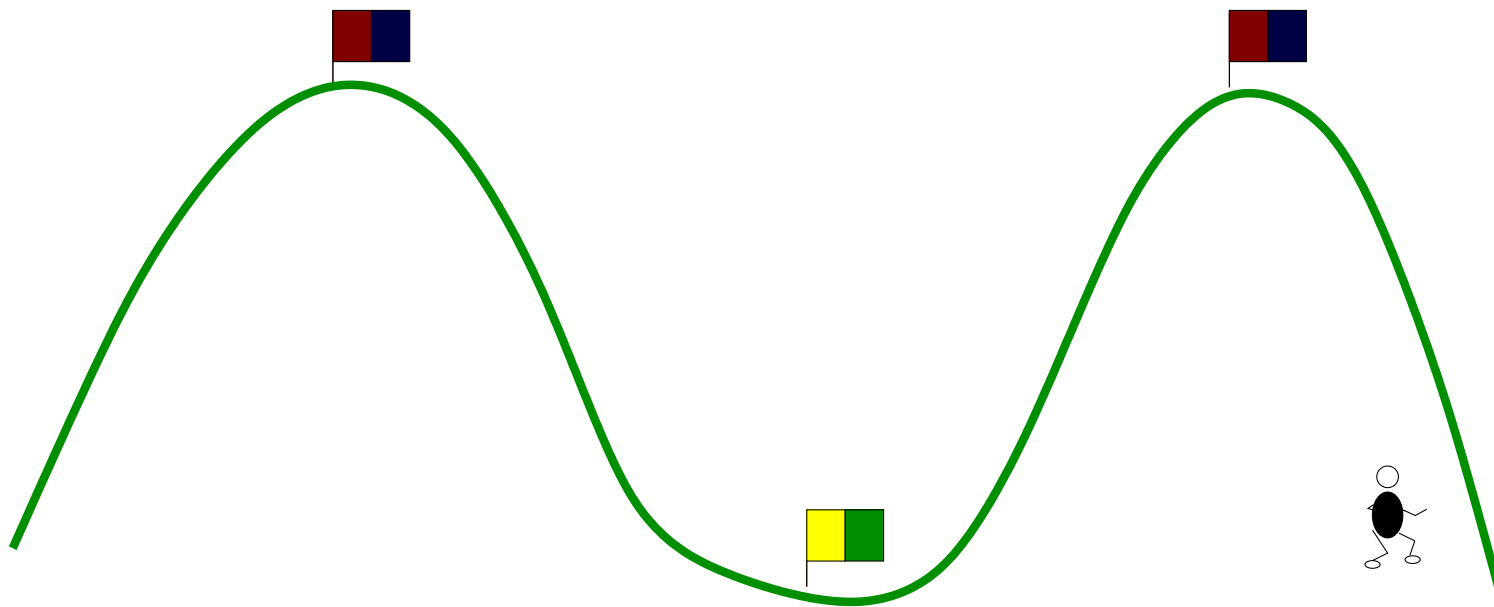
***But, the messenger can
get lost !***

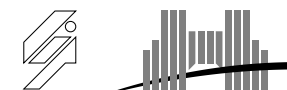
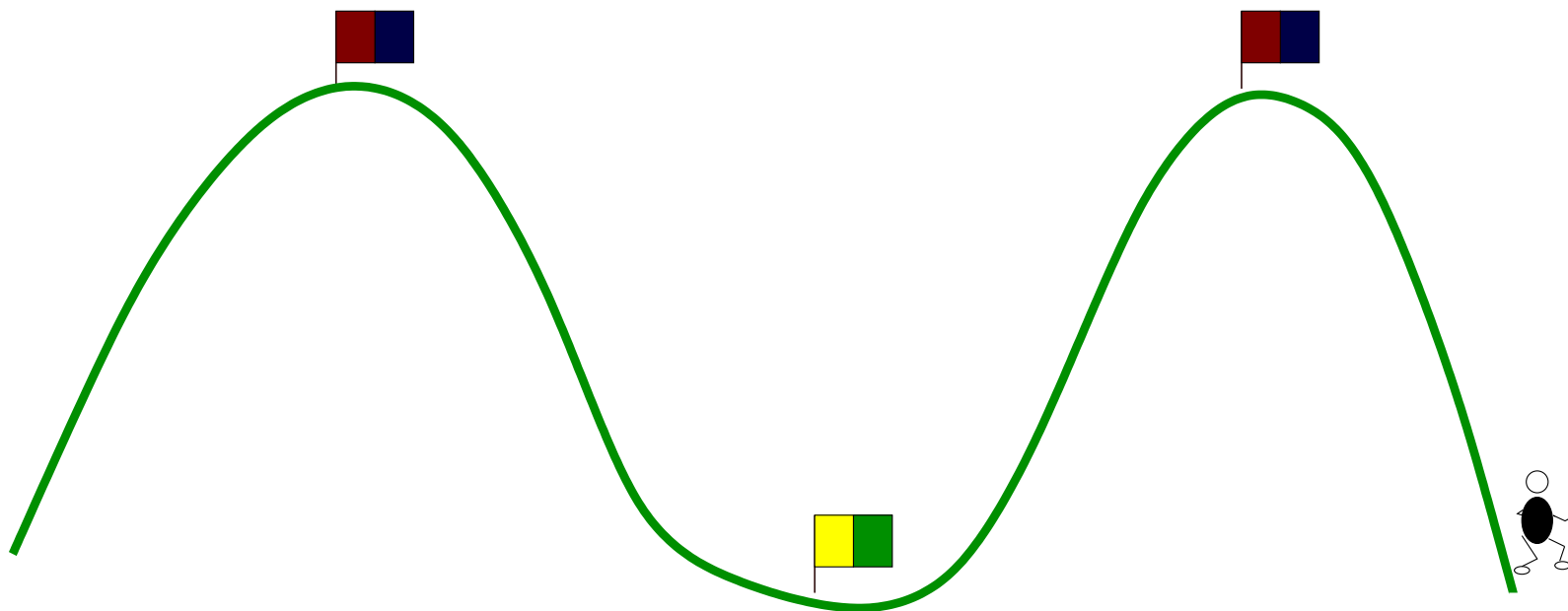












The coordinated attack

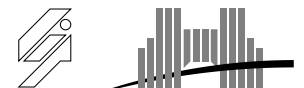
General 1 chooses a time for the attack, say H , and sends a messenger.

Upon arrival of the messenger, general 2 agrees on the hour H and sends a messenger with an agreement.

General 1 will attack at time H if he knows that General 2 knows his proposed hour and agrees on.

General 2 will attack at time H if he (General 2) knows that General 1 knows that he (General 2) knows the proposed hour H .

General 1 must send a second messenger with an acknowledgment.



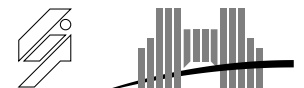
General 1 will attack at time H if he (General 1) knows that General 2 knows that he (General 1) knows that General 2 knows the proposed hour.

General 2 must send a second messenger with an acknowledgment.

General 2 will attack at time H if he (General 2) knows that General 1 knows that he (General 2) knows that General 1 knows that he (General 2) knows the proposed hour H .

General 1 must send a third messenger with an acknowledgment.

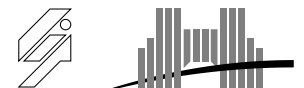
⋮



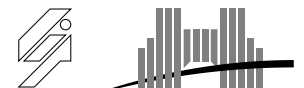
The coordinated attack

The process goes forever.

One can prove that, with asynchronized communications,
a coordinated attack is **not** possible.



Security on Internet



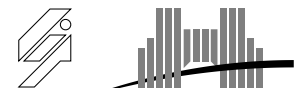
Security on Internet

The goal is to **transform** sentences “**I believe that ...**”
into sentences “**I know that ...**”.

Messages are encoded and traverse a public network,
but this is not enough.

Intruders on the network can

- listen to messages,
- stock them
- and replay them or build fake messages.

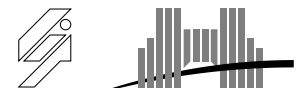


Security on Internet

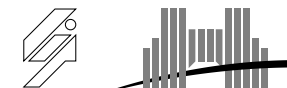
Assume A received a message from B .

A must be able to assert

***"I know that** the message I received has been sent by B ".*



The logic of knowledge

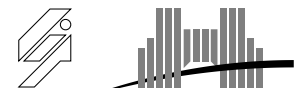


The modalities

A modality is an operator which **transforms** a sentence in another sentence.

One creates a modality K_A for each agent A .

A logic with modalities is called a **modal logic**.

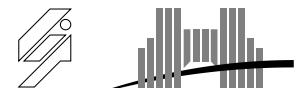


What is logic of knowledge ?

The logic of knowledge or epistemic logic

is the logic that formalizes

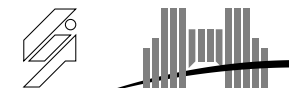
- “the agent i knows that φ ”, written $K_i(\varphi)$,
- “ φ is a common knowledge”, written $C_G(\varphi)$.



Common knowledge

$C_G(\varphi)$ formalizes sentences like

- “It is a well known fact that φ , except for mad people.”
- “Agent i knows that agent j knows that agent i knows that agent j knows that, etc.”.



Common knowledge

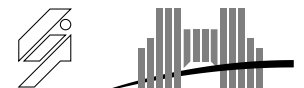
$C_G(\varphi)$ formalizes sentences like

- “It is a well known fact that φ , except mad people.”
- “Agent i knows that agent j knows that agent i knows that agent j knows that, etc.”.

One needs a modality E , called “shared knowledge”, that says

“Everybody knows that φ ”

$$E_G(\varphi) = \bigwedge_{i \in G} K_i(\varphi).$$



Common knowledge

$C_G(\varphi)$ formalizes sentences like

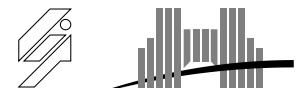
- “It is a well known fact that φ , except mad people.”
- “Agent i knows that agent j knows that agent i knows that agent j knows that, etc.”.

One needs a modality E , called “shared knowledge”, that says

“Everybody knows that φ ”,

$$E_G(\varphi) = \bigwedge_{i \in G} K_i(\varphi).$$

Common knowledge is not shared knowledge.



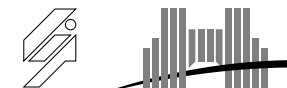
Common knowledge

$C_G(\varphi)$ is the fixpoint of

$$\psi \Leftrightarrow \varphi \wedge E_G(\psi)$$

i.e.,

$$C_G(\varphi) \Leftrightarrow \varphi \wedge E_G(C_G(\varphi))$$



Hi, who are you ?

Am Napoleon.

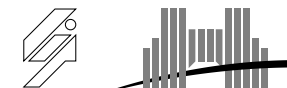
Yeah. Who told you that?

God told me.

Did I say that?



Rules and axioms



Rules

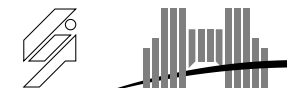
It is a logic à la Hilbert.

Modus ponens

$$\frac{\vdash \varphi \quad \vdash \varphi \Rightarrow \psi}{\vdash \psi} \text{ (MP)}$$

Knowledge generalization

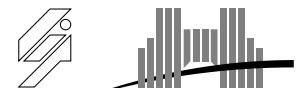
$$\frac{\vdash \varphi}{\vdash K_i(\varphi)} \text{ (KG)}$$



The axioms

All theorems of traditional logic.

$\frac{}{\vdash \varphi}$ (CI) if φ is a theorem of logic.



The axioms

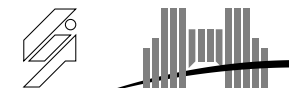
Four more axioms

Distribution axiom

$$\frac{}{\vdash K_i(\varphi) \Rightarrow K_i(\varphi \Rightarrow \psi) \Rightarrow K_i(\psi)} \text{ (K)}$$

Knowledge axiom

$$\frac{}{\vdash K_i(\varphi) \Rightarrow \varphi} \text{ (T)}$$



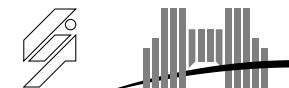
The axioms

Positive introspection axiom

$$\frac{}{\vdash K_i(\varphi) \Rightarrow K_i(K_i(\varphi))} \quad (4)$$

Negative introspection axiom

$$\frac{}{\vdash \neg K_i(\varphi) \Rightarrow K_i(\neg K_i(\varphi))} \quad (5)$$



Beware

In modal logic, one cannot take plain natural deduction.

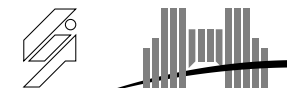
One can use “natural sequents” like $\Gamma \vdash \varphi$.

But the **knowledge generalization** gives

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

where $K_i(\Gamma)$ means that one puts a K_i in front of all the propositions in Γ .

The operation $K_i(\Gamma)$ is not a traditional operation in natural deduction.



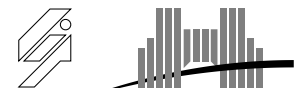
The axioms of common knowledge

Definition of E_G

$$\frac{}{\vdash E_G(\varphi) \Leftrightarrow \bigwedge_{i \in G} K_i(\varphi)} \quad (C1)$$

$C_G(\varphi)$ satisfies the inequality $\psi \Rightarrow \varphi \wedge E_G(\psi)$.

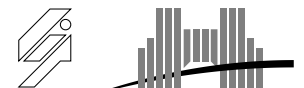
$$\frac{}{\vdash C_G(\varphi) \Rightarrow \varphi \wedge E_G(C_G(\varphi))} \quad (C2)$$



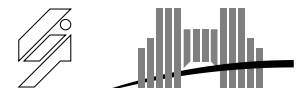
The rules of common knowledge

$C_G(\varphi)$ is **the least** in some sense, that is
if any ψ satisfies $\psi \Rightarrow \varphi \wedge E_G(\psi)$
then $\psi \Rightarrow C_G(\varphi)$.

$$\frac{\vdash \psi \Rightarrow \varphi \wedge E_G(\psi)}{\vdash \psi \Rightarrow C_G(\varphi)} \quad (RC1)$$



The models



The Kripke models

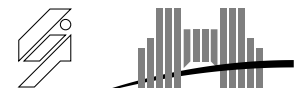
A **Kripke model** is a triple $\mathcal{M} = (\mathcal{U}_{\mathcal{M}}, \mathcal{I}_{\mathcal{M}}, \mathcal{R}_{\mathcal{M}})$ where

- $\mathcal{U}_{\mathcal{M}}$ is a set of elements which are called **worlds**,
- $\mathcal{I}_{\mathcal{M}} : \text{Variables} \rightarrow \mathcal{P}(\mathcal{U}_{\mathcal{M}})$.
Intuitively $\mathcal{I}_{\mathcal{M}}(p)$ is the set of worlds where agent i knows that variable p is satisfied.
- $\mathcal{R}_{\mathcal{M}} = (R_1, \dots, R_n)$ is a set of equivalence relations (one by agent) called **accessibility relations**.

If $u R_i v$ then the world v is **accessible** from u for i .

If $\mathcal{I}_{\mathcal{M}}(p)$ contains a world u ,

then it must contain all the words v such that $u R_i v$ for all i .



A simile game

2 agents, 3 cards $\{A, B, C\}$.

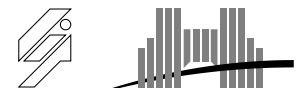
Agent 1 receives one card

Agent 2 receives one card

The third card is face down.

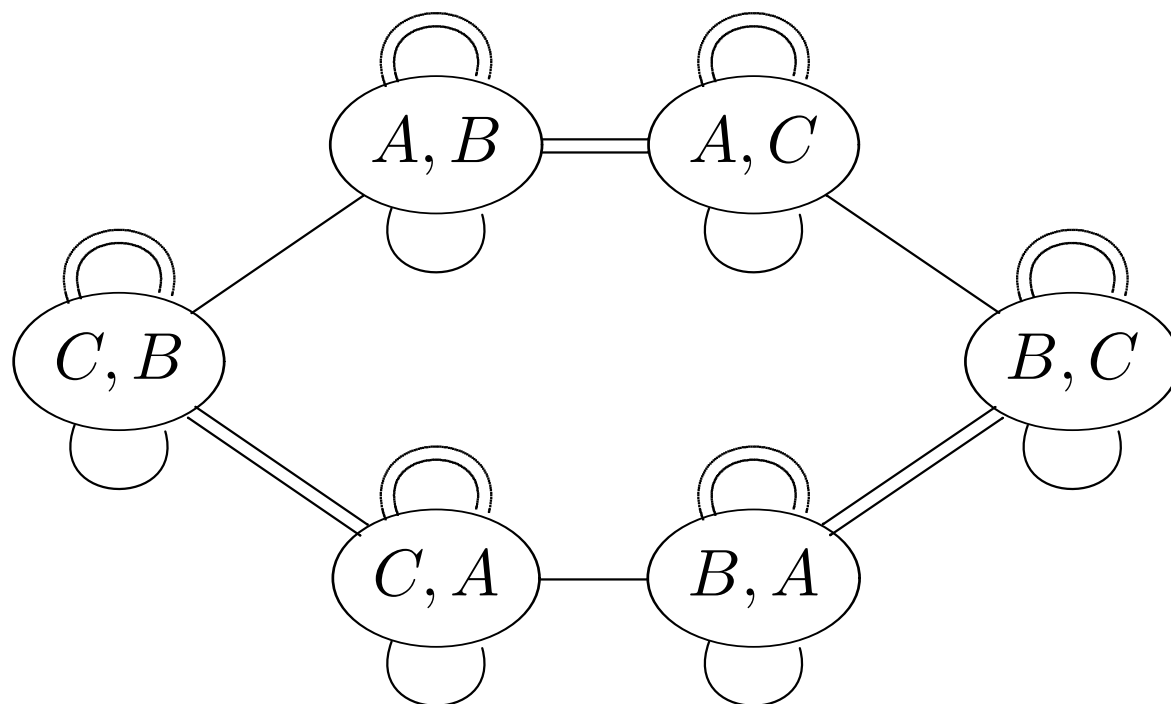
There are six possible worlds :

$(A, B), (A, C), (B, A), (B, C), (C, A), (C, B)$.



A simple game

In worlds (A, B) agent 1 (its accessibility relation is written \equiv) accepts two possible worlds namely (A, B) and (A, C) .

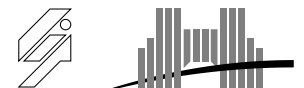


The Kripke model \mathcal{M} .

A simple game

Primitive propositions are

- $1A$ player (agent) 1 holds card A ,
- $2A$ player (agent) 2 holds card A ,
- $1B$ player (agent) 1 holds card B ,
- $2B$ player (agent) 2 holds card B ,
- $1C$ player (agent) 1 holds card C ,
- $2C$ player (agent) 2 holds card C .



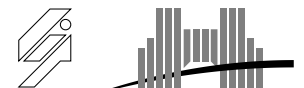
Some forcing assertions

$$(A, B) \Vdash 1A \wedge 2B,$$

$$(A, B) \Vdash K_1(2B \vee 2C),$$

$$(A, B) \Vdash K_1(\neg K_2(1A)).$$

For all worlds u the assertion $u \Vdash K_1(2A \vee 2B \vee 2C)$ holds
hence $\mathcal{M} \models K_1(2A \vee 2B \vee 2C)$.



Accessibility and forcing

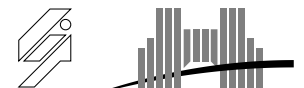
One must have

$$u \Vdash K_i \varphi \iff (\forall v \in \mathcal{U}_{\mathcal{M}}) v R_i u \Rightarrow v \Vdash \varphi.$$

This means also that

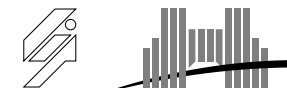
agent i knows φ in world u

if and only if in each worlds that he takes as possible, φ holds.



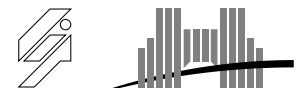
The puzzle of the muddy children

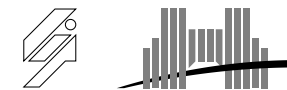
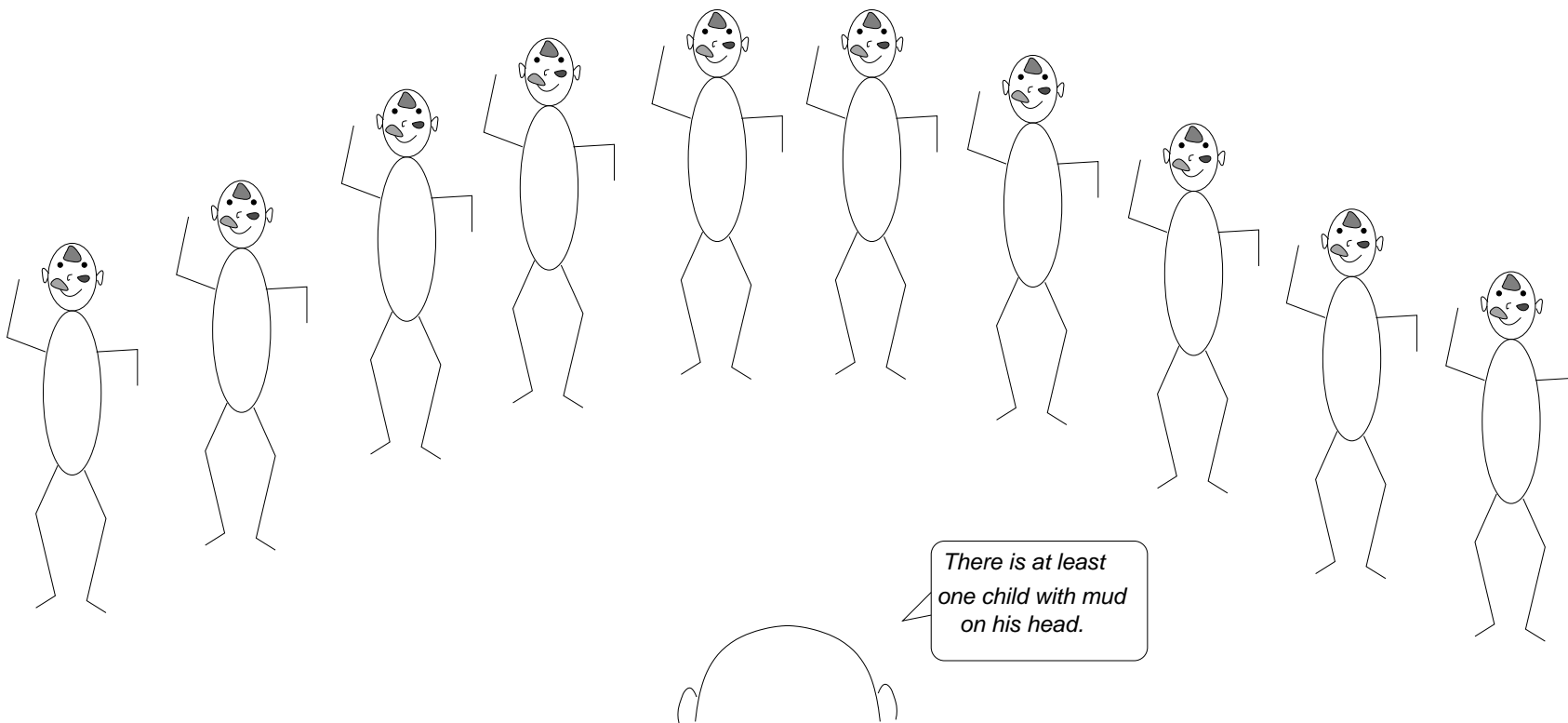
March 10, 2004



The muddy children

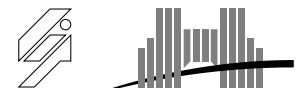
- There are n children some of them have mud on their head.
- Father says “There is at least one child with mud on his head”.



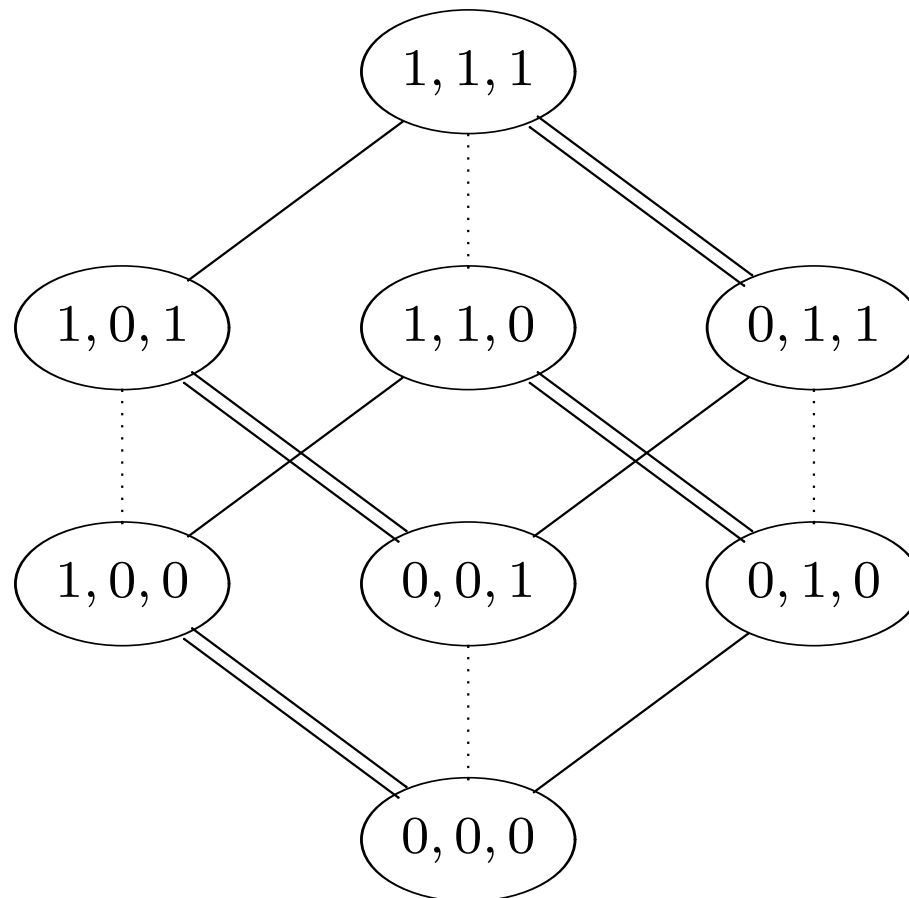


The muddy children

- There are n children some of them have mud on their head.
- Father says “There is at least one child with mud on his head”.
- Then Father tells many times (how many ?) the following request
“If you have mud on your head, please step forward.”.
- As n children have mud on their head,
- after n requests, they all step forward.

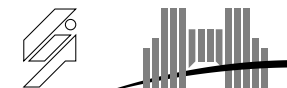
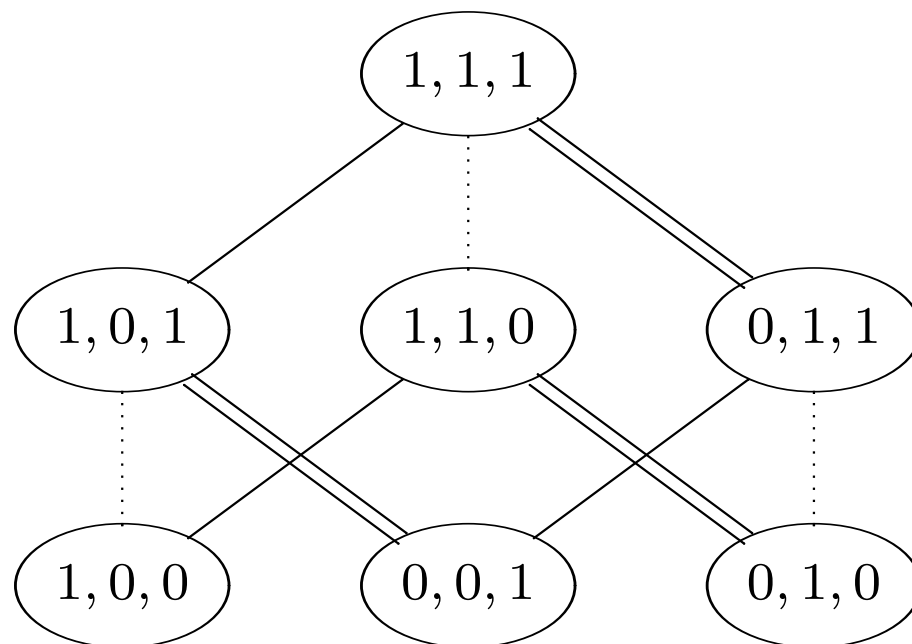


Kripke model for three muddy children

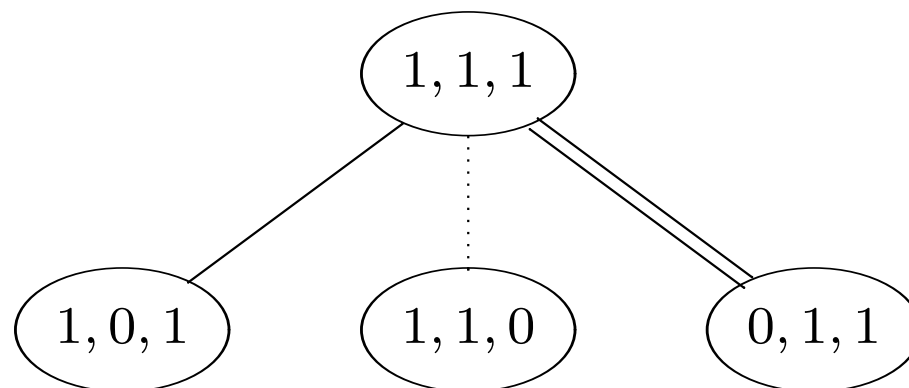


One drops reflexivity loops.

After Father has spoken

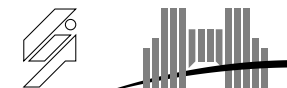


After Father has asked his first request

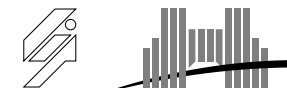


After Father has asked his second request

1, 1, 1

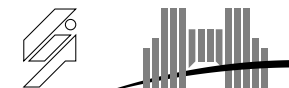


Correction and proofs



Correction

Theorem : If $\vdash \varphi$ then $\models \varphi$.



Why not a deduction rule ?

If one takes the **deduction rule**

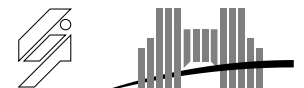
“From $\Gamma, \varphi \vdash \psi$ deduce $\Gamma \vdash \varphi \Rightarrow \psi$ ”

then from the judgment $\varphi \vdash K_i(\varphi)$ one would get $\varphi \vDash K_i(\varphi)$,

that is “If in all the world of the universe, φ holds,
then each agent i knows φ ”

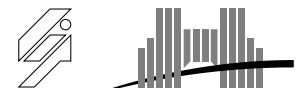
one could deduce $\vDash \varphi \Rightarrow K_i(\varphi)$

that is “If φ holds then each agent i knows φ ”.



A proof

One can prove $\vdash \varphi \Rightarrow K_i(\neg K_i(\neg \varphi))$.



A proof

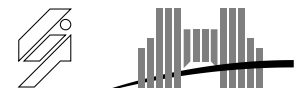
One can prove $\vdash \varphi \Rightarrow K_i(\neg K_i(\neg\varphi))$.

$$\begin{array}{c}
 \frac{}{\vdash \psi} \text{ (CI)} \qquad \frac{}{\vdash K_i(\neg\varphi) \Rightarrow \neg\varphi} \text{ (T)} \\
 \hline
 \vdash \neg K_i(\varphi) \Rightarrow K_i(\neg K_i(\neg\varphi)) \text{ (5)} \qquad \vdash (\neg K_i(\varphi) \Rightarrow K_i(\neg K_i(\neg\varphi))) \Rightarrow \varphi \Rightarrow K_i(\neg K_i(\neg\varphi)) \text{ (MP)} \\
 \hline
 \vdash \varphi \Rightarrow K_i(\neg K_i(\neg\varphi)) \text{ (MP)}
 \end{array}$$

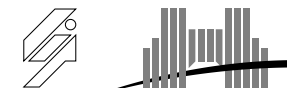
where $\psi \equiv (K_i(\neg\varphi) \Rightarrow \neg\varphi) \Rightarrow (\neg K_i(\varphi) \Rightarrow K_i\neg K_i\neg\varphi) \Rightarrow \varphi \Rightarrow K_i(\neg K_i(\neg\varphi))$

which is a classic theorem.

Indeed this is an instance of $(B \Rightarrow \neg A) \Rightarrow (\neg B \Rightarrow C) \Rightarrow (A \Rightarrow C)$.



The COQ formalization



The type of propositions

A proposition is either

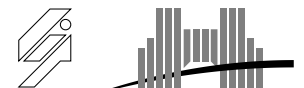
- an implication,
- or a universal quantification,
- or a modal “knowledge” proposition with a K ,
- or a modal “common knowledge” proposition with a C .

Inductive proposition: Set :=

```

Imp      :  proposition -> proposition -> proposition |
forall   :  (A:Set) (A -> proposition) -> proposition |
K        :  nat -> proposition -> proposition          |
C        :  (list nat) -> proposition -> proposition.

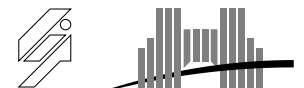
```



Agent as natural

Agents are represented by natural numbers.

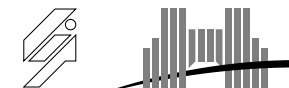
Groups of agents are lists of naturals.



The meta-predicate theorem

`theorem` tells which propositions are theorems.

For instance, $(\text{theorem } p)$ says that proposition p is a theorem in the object theory representing epistemic logic.



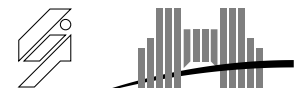
Axioms and rules

Axioms are just given by declaring basic theorems.

Hilbert_K: $(p, q: \text{proposition}) \text{ (theorem } p \Rightarrow q \Rightarrow p)$

Hilbert_S: $(p, q, r: \text{proposition})$
 $\text{(theorem } (p \Rightarrow q \Rightarrow r) \Rightarrow (p \Rightarrow q) \Rightarrow p \Rightarrow r)$

MP: $(p, q: \text{proposition}) \text{ (theorem } p \Rightarrow q) \rightarrow \text{(theorem } p)$
 $\rightarrow \text{(theorem } q).$

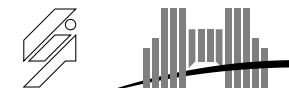


Axioms and rules

Hilbert_K: (p,q:proposition) (theorem p => q => p)

should be read

$$(\forall p, q \in \textit{proposition}) \vdash p \Rightarrow q \Rightarrow p$$



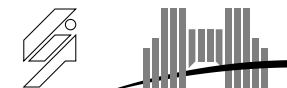
Axioms and rules

Hilbert_S: (p,q,r:proposition)

(theorem (p => q => r) => (p => q) => p => r)

should be read

$(\forall p, q, r \in \text{proposition}) \vdash (p \Rightarrow q \Rightarrow r) \Rightarrow (p \Rightarrow q) \Rightarrow (p \Rightarrow r)$



Axioms and rules

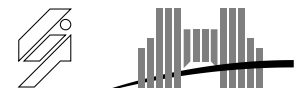
MP: $(p, q: \text{proposition}) \text{ (theorem } p \Rightarrow q) \rightarrow \text{(theorem } p) \rightarrow \text{(theorem } q)$.

should be read

$(\forall p, q, r \in \text{proposition})$ if $\vdash (p \Rightarrow q)$ and $\vdash p$ then $\vdash q$.

which can be written

$$(\forall p, q, r \in \text{proposition}) \frac{\vdash p \quad \vdash p \Rightarrow q}{\vdash q}$$



The proof

The proofs require using only Hilbert proofs.

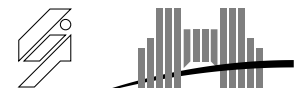
For that one uses systematically the **Cut Rule**

$(p, q, r : \text{proposition})$

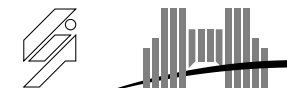
$(\text{theorem } p \Rightarrow q) \rightarrow (\text{theorem } q \Rightarrow r) \rightarrow (\text{theorem } p \Rightarrow r).$

which is

$$\frac{\vdash p \Rightarrow q \quad \vdash q \Rightarrow r}{\vdash p \Rightarrow r}$$

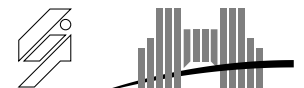


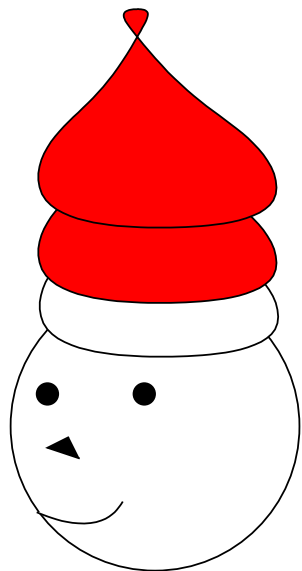
The king, the three wise men and the hats



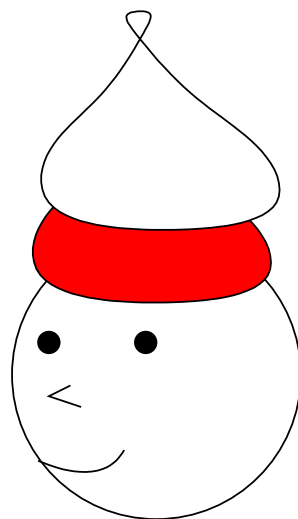
The statement

“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”

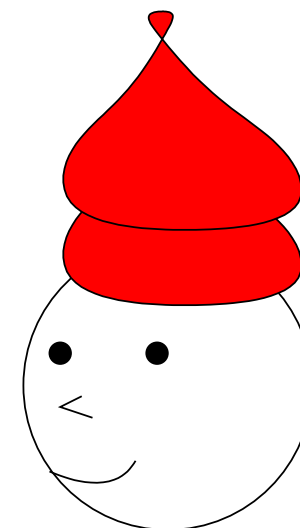




Alice



Bob



Carol

A definition and the main theorem

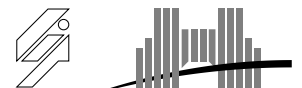
An agent knows the color of his (her) hat.

Definition $Kh := [i:\text{nat}] (K\ i\ (\text{white}\ i)) \mid / (K\ i\ (\text{red}\ i))$.

With a minimal set of hypotheses, we are able to prove

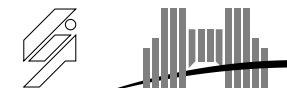
(theorem

$(K\ \text{Bob}\ (\text{Not}\ (Kh\ \text{Alice}))) \ \&\ (\text{Not}\ (Kh\ \text{Bob})) \Rightarrow (\text{red}\ \text{Carol})$).



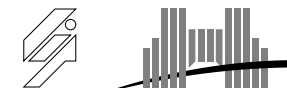
The proof requires only modal logic.

There is no common knowledge.



The puzzle of the muddy children

March 10, 2004



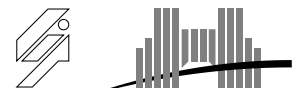
Two predicates

At_least and Exactly

$(\text{At_least } n \ p)$ is intended to mean that among the n children, there are at least p muddy children.

Exactly means that among the n children, there are exactly p muddy children.

Exactly is defined as

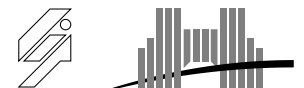
$$[n, p : \text{nat}] (\text{At_least } n \ p) \ \& \ (\text{Not } (\text{At_least } n \ (S \ p))).$$


The axiom of Knowledge diffusion

```

Axiom Knowledge_Diffusion : (n,p,i:nat)
  (theorem (E (list_of n) (At_least n p))
    => (E (list_of n) (Not (Exactly n p))))
    => (K i (E (list_of n) (Not (Exactly n p))))).
  
```

$$\begin{aligned}
 \vdash E_{\text{Chd}_n}(At_least(n,p)) &\Rightarrow E_{\text{Chd}_n}(\neg Exactly(n,p)) \\
 &\Rightarrow K_i(E_{\text{Chd}_n}(\neg Exactly(n,p))).
 \end{aligned}$$



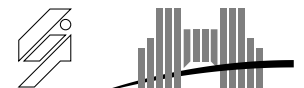
Two theorems

```

Lemma E_Awareness : (n,p:nat)
  (theorem (E (list_of n) (At_least n p))
    => (E (list_of n) (Not (Exactly n p))))
  => (E (list_of n) (E (list_of n) (Not (Exactly n p)))).

```

$$\begin{aligned}
 \vdash E_{\text{Chd}_n}(\textit{At_least}(n, p)) &\Rightarrow E_{\text{Chd}_n}(\neg \textit{Exactly}(n, p)) \\
 &\Rightarrow E_{\text{Chd}_n}(E_{\text{Chd}_n}(\neg \textit{Exactly}(n, p)))
 \end{aligned}$$

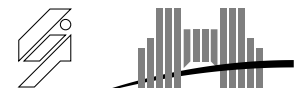


Two theorems (next)

Lemma `C_Awareness` : $(n, p : \text{nat})$
 $(\text{theorem } (C \text{ (list_of } (S \ n)) \text{ (At_least } (S \ n) \ p))$
 $\Rightarrow (E \text{ (list_of } (S \ n)) \text{ (Not } ((\text{Exactly } (S \ n) \ p))))$
 $\Rightarrow ((C \text{ (list_of } (S \ n)) \text{ (Not } (\text{Exactly } (S \ n) \ p))))).$

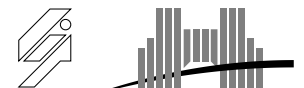
$$\begin{aligned} \vdash C_{\text{Chd}_{n+1}}(\text{At_least}(n+1, p)) &\Rightarrow E_{\text{Chd}_{n+1}}(\neg \text{Exactly}(n+1, p)) \\ &\Rightarrow C_{\text{Chd}_{n+1}}(E_{\text{Chd}_{n+1}}(\neg \text{Exactly}(n+1, p))) \end{aligned}$$

C_Awareness can only be proved for a non empty group of children.

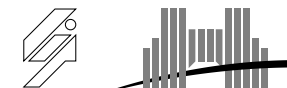


The main theorem

$$\begin{aligned}
 & (C \text{ (list_of (S n)) (At_least (S n) p)}) \\
 & \& (E \text{ (list_of (S n)) (Not (Exactly (S n) p))}) \\
 & \Rightarrow (C \text{ (list_of (S n)) (At_least (S n) (S p))}).
 \end{aligned}$$

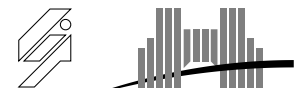
$$\begin{aligned}
 \vdash C_{\text{Chd}_{n+1}}(\text{At_least}(n+1, p)) & \Rightarrow E_{\text{Chd}_{n+1}}(\neg \text{Exactly}(n+1, p)) \\
 & \Rightarrow C_{\text{Chd}_{n+1}}(\text{At_least}(n+1, p+1))
 \end{aligned}$$


What we learned



“Books are usually wrong.”

- Toelstra and van Dalen give a wrong axiomatization of Forall .
- Fagin et al. on one hand and Meyer and van der Hoek on the other hand give a wrong claim about common knowledge in the case of an empty group.



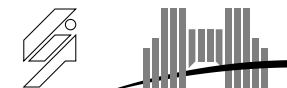
The organization

- Notations
- Version control

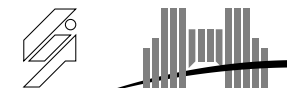
The proof

- Management of proofs à la Hilbert
- Acceptable hypotheses

A didactic tool



That's all !



He **believes** he is Napoleon,
but **it is well known**
that I am Napoleon.

