

# “Cognitive” Deductive Shots @ *R*

*Are Humans Rational?*  
RPI

Selmer Bringsjord & Atriya Sen

9.12.19

# “Cognitive” Attack on $\mathcal{R}$

\*Makes use of such mental verbs as *believing*, *knowing*, etc.

# “Cognitive” Attack on $\mathcal{R}$

If humans are as described in this thesis, then they can solve the forthcoming “cognitive”\* problems.

\*Makes use of such mental verbs as *believing*, *knowing*, etc.

# “Cognitive” Attack on $\mathcal{R}$

If humans are as described in this thesis, then they can solve the forthcoming “cognitive”\* problems.

But humans *can't* solve the problems in question.

\*Makes use of such mental verbs as *believing*, *knowing*, etc.

# “Cognitive” Attack on $\mathcal{R}$

If humans are as described in this thesis, then they can solve the forthcoming “cognitive”\* problems.

But humans *can't* solve the problems in question.

Therefore:

\*Makes use of such mental verbs as *believing*, *knowing*, etc.

# “Cognitive” Attack on $\mathcal{R}$

If humans are as described in this thesis, then they can solve the forthcoming “cognitive”\* problems.

But humans *can't* solve the problems in question.

Therefore:

Sorry Selmer & company, your thesis  $\mathcal{R}$  is false.

\*Makes use of such mental verbs as *believing*, *knowing*, etc.

# “Cognitive” Attack on $\mathcal{R}$

If humans are as described in this thesis, then they can solve the forthcoming “cognitive”\* problems.

But humans *can't* solve the problems in question.

Therefore:

Sorry Selmer & company, your thesis  $\mathcal{R}$  is false.

\*Makes use of such mental verbs as *believing*, *knowing*, etc.



# Floridi's Continuum (augmented), and Claims

(“Consciousness, Agents, and the Knowledge Game” *Minds & Machines*)

	False Belief Task	Wise Man Test ( <i>n</i> )	Deafening Test	Torture Boots Test	Ultimate Sifter	<i>Infinitary</i> False Belief Task
Cutting-Edge AI	Yes	Yes	No	No	No	?
Zombies	Yes	Yes	Yes	Yes	No	?
Human Persons (s-conscious! p-conscious!)	Yes	Yes	Yes	Yes	Yes	Yes



# Cracking False-Belief Tasks ...



# In SL, w/ real-time comm w/ ATP





# In SL, w/ real-time comm w/ ATP





```
SNARK-USER 14 >
(in-immature-scenario
  (prove '(t-retrieve subject
                    teddybear
                    ?c)
    :answer '(looks-in ?c)))
```

(Refutation

```
(Row 1
  (or (not (person ?x)) (not (object ?y))
    (not (container ?z)) (not (in ?y ?z))
    (bel-in ?x ?y ?z))
  assertion)
(Row 2
  (or (not (person ?x))
    (not (container ?y))
    (not (object ?z))
    (not (w-retrieve ?x ?z))
    (not (bel-in ?x ?z ?y))
    (t-retrieve ?x ?z ?y))
  assertion)
(Row 4
  (person subject)
  assertion)
(Row 6
  (container c2)
  assertion)
(Row 7
  (object teddybear)
  assertion)
(Row 8
```

```
    (in teddybear c2)
    assertion)
(Row 9
  (w-retrieve subject teddybear)
  assertion)
(Row 10
  (not (t-retrieve subject teddybear ?x))
  negated_conjecture
  Answer (looks-in ?x))
(Row 11
  (or (not (person ?x)) (bel-in ?x
teddybear c2))
  (rewrite (resolve 1 8) 6 7))
(Row 25
  (bel-in subject teddybear c2)
  (resolve 11 4))
(Row 28
  (t-retrieve subject teddybear c2)
  (rewrite (resolve 2 25) 9 7 6 4))
(Row 30
  false
  (resolve 10 28)
  Answer (looks-in c2)))

:PROOF-FOUND

SNARK-USER 15 > (answer t)
(LOOKS-IN C2)
```



```
SNARK-USER 12 >
(in-mature-scenario
  (prove '(t-retrieve subject
                    teddybear
                    ?c)
    :answer '(looks-in ?c)))
```

```
(Refutation
(Row 1
  (or (not (person ?x))
      (not (container ?y))
      (not (object ?z))
      (not (w-retrieve ?x ?z))
      (not (bel-in ?x ?z ?y))
      (t-retrieve ?x ?z ?y))
  assertion)
(Row 2
  (or (not (person ?x)) (not (object ?
y)) (not (container ?z)) (not (p-in ?x ?y
?z)) (bel-in ?x ?y ?z))
  assertion)
(Row 4
  (person subject)
  assertion)
(Row 5
  (container c1)
  assertion)
(Row 7
  (object teddybear)
  assertion)
(Row 8
  (p-in subject teddybear c1)
```

```
  assertion)
(Row 9
  (w-retrieve subject teddybear)
  assertion)
(Row 10
  (not (t-retrieve subject teddybear ?
x))
  negated_conjecture
  Answer (looks-in ?x))
(Row 11
  (bel-in subject teddybear c1)
  (rewrite (resolve 2 8) 5 7 4))
(Row 25
  (t-retrieve subject teddybear c1)
  (rewrite (resolve 1 11) 9 7 5 4))
(Row 26
  false
  (resolve 10 25)
  Answer (looks-in c1))
)

:PROOF-FOUND

SNARK-USER 13 > (answer t)
(LOOKS-IN C1)
```



“The present account of the false belief transition is incomplete in important ways. After all, our agent had only to choose the best of two known models. This begs an understanding of the dynamics of rational revision near threshold and when the space of possible models is far larger. Further, a single formal model ought ultimately to be applicable to many false belief tasks, and to reasoning about mental states more generally. Several components seem necessary to extend a particular theory of mind into such a framework theory: a richer representation for the propositional content and attitudes in these tasks, extension of the implicit quantifier over trials to one over situations and people, and a broader view of the probability distributions relating mental state variables. Each of these is an important direction for future research.”

“Intuitive Theories of Mind: A Rational Approach to False Belief”  
Goodman et al.



“The present account of the false belief transition is incomplete in important ways. After all, our agent had only to choose the best of two known models. This begs an understanding of the dynamics of rational revision near threshold and when the space of possible models is far larger. Further, a single formal model ought ultimately to be applicable to many false belief tasks, and to reasoning about mental states more generally. Several components seem necessary to extend a particular theory of mind into such a framework theory: a richer representation for the propositional content and attitudes in these tasks, extension of the implicit quantifier over trials to one over situations and people, and a broader view of the probability distributions relating mental state variables. Each of these is an important direction for future research.”

“Intuitive Theories of Mind: A Rational Approach to False Belief”  
Goodman et al.



Done.

“The present account of the false belief transition is incomplete in important ways. After all, our agent had only to choose the best of two known models. This begs an understanding of the dynamics of rational revision near threshold and when the space of possible models is far larger. Further, a single formal model ought ultimately to be applicable to many false belief tasks, and to reasoning about mental states more generally. Several components seem necessary to extend a particular theory of mind into such a framework theory: a richer representation for the propositional content and attitudes in these tasks, extension of the implicit quantifier over trials to one over situations and people, and a broader view of the probability distributions relating mental state variables. Each of these is an important direction for future research.”

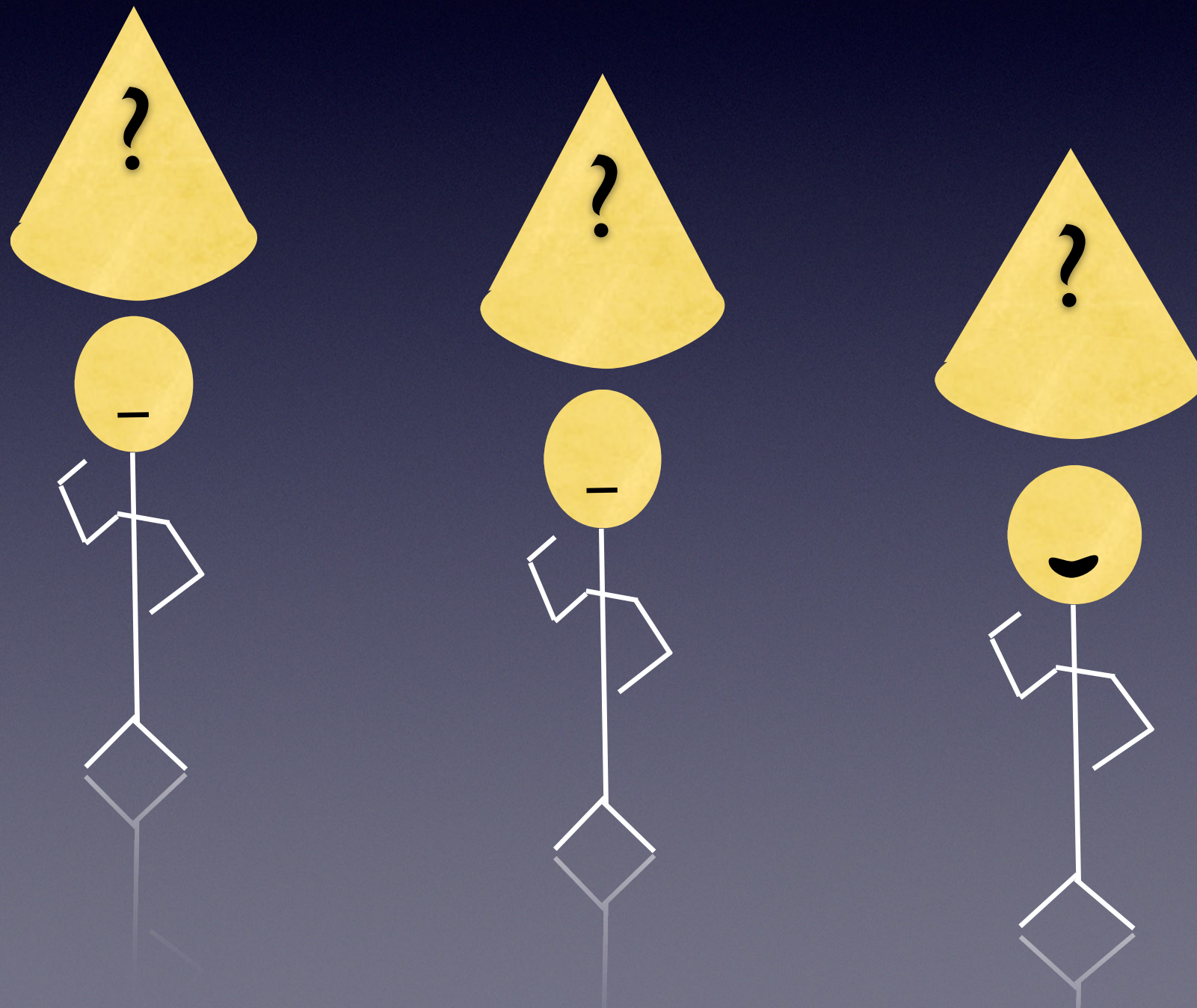
Done.

“Intuitive Theories of Mind: A Rational Approach to False Belief”  
Goodman et al.



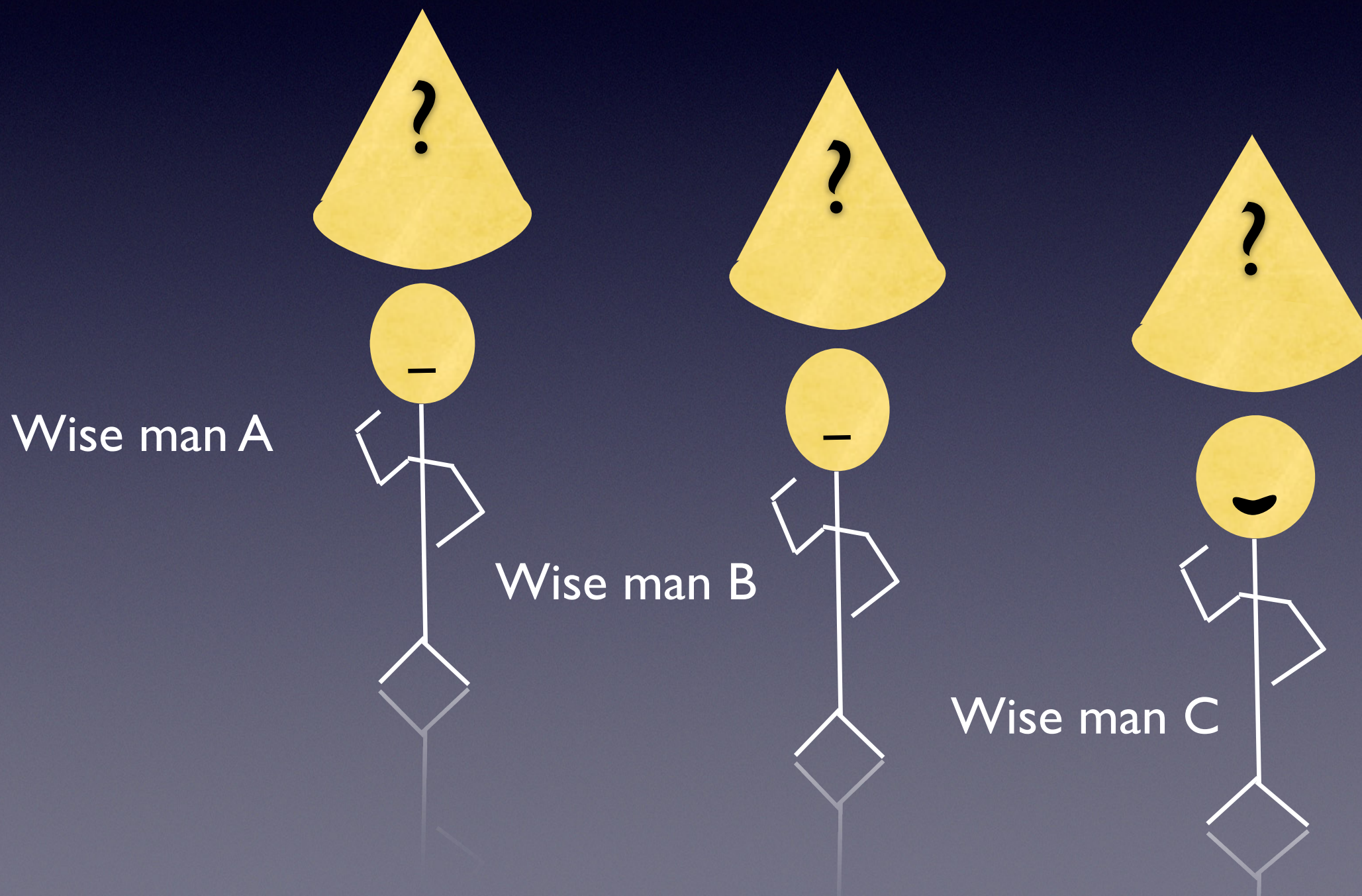
Cracking Wise Man Tests ...

# Wise Men Puzzle





# Wise Men Puzzle





# Wise Men Puzzle

I don't know

?

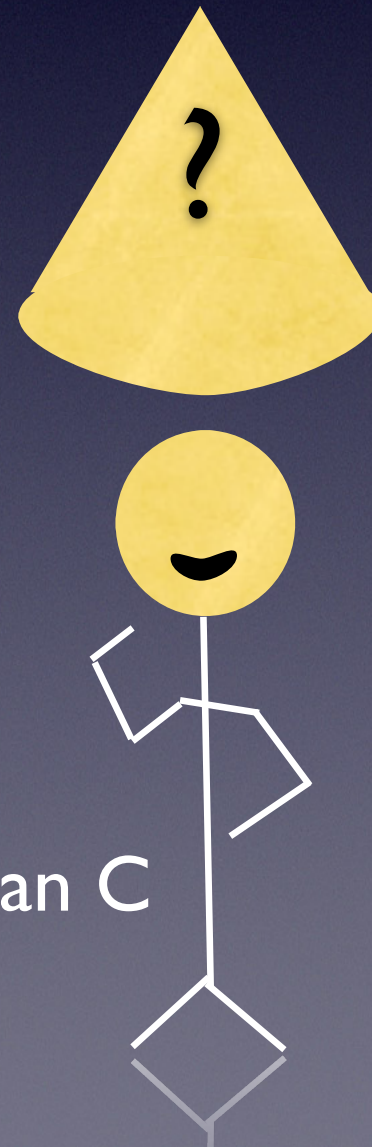
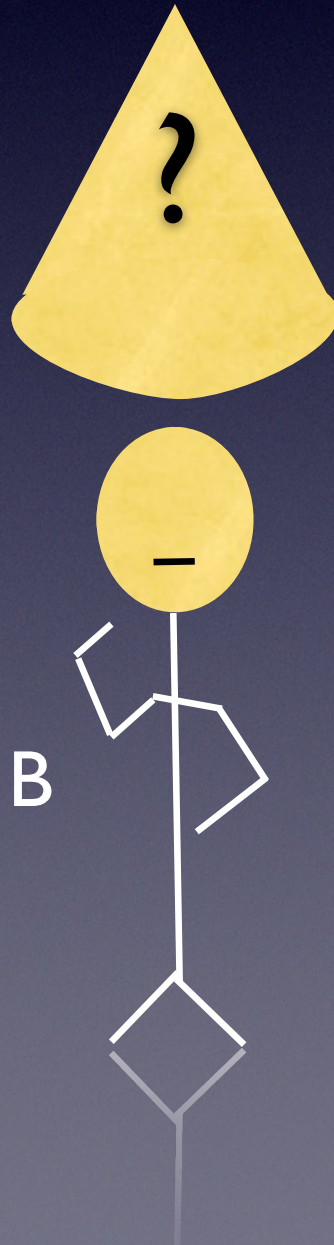
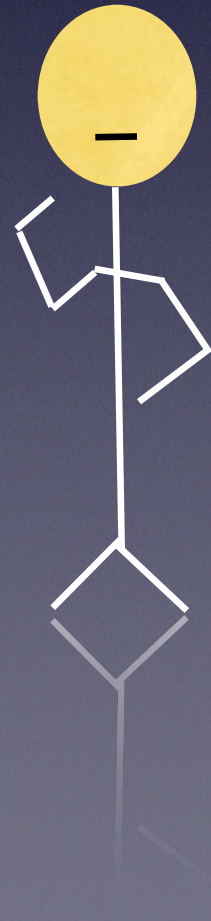
?

?

Wise man A

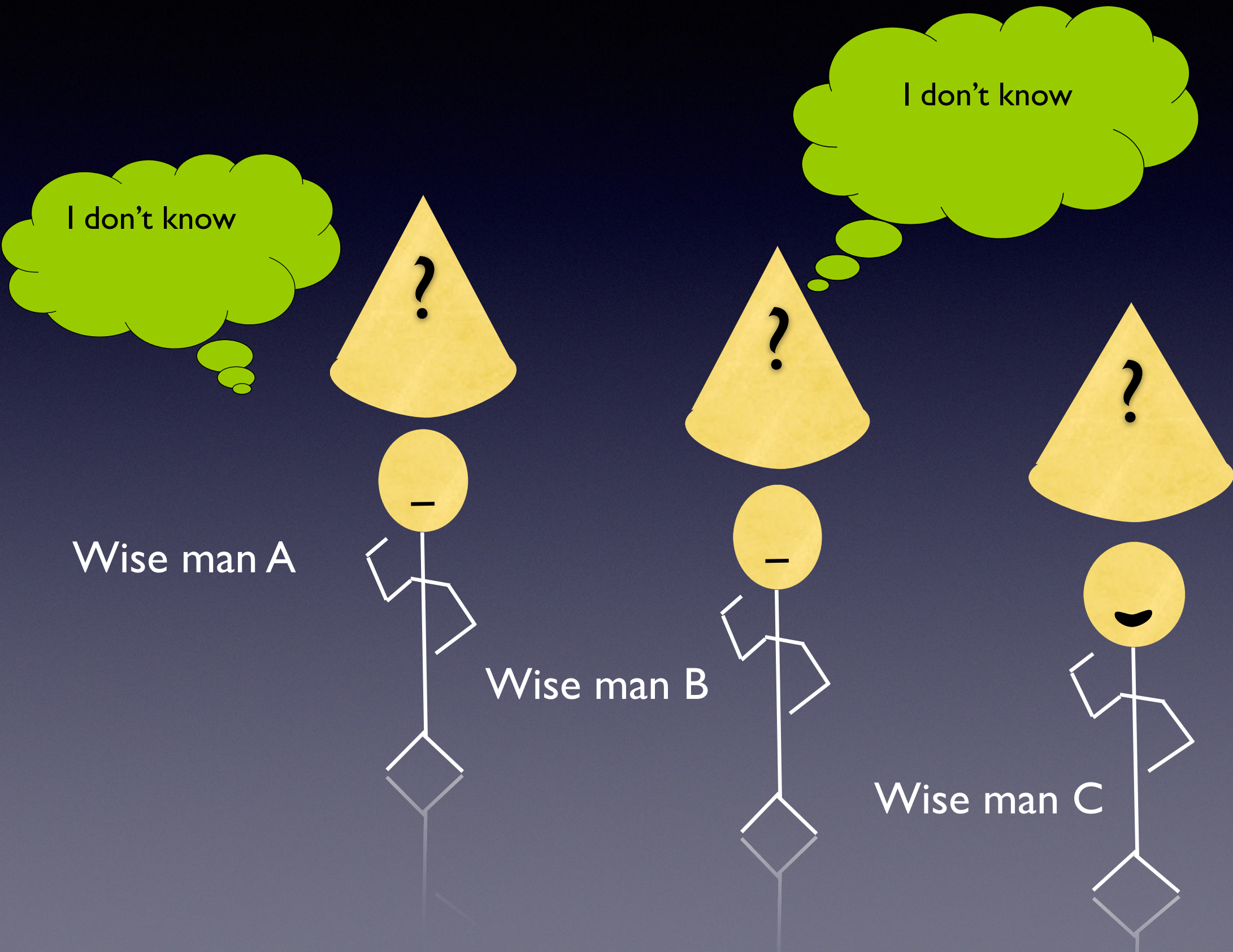
Wise man B

Wise man C



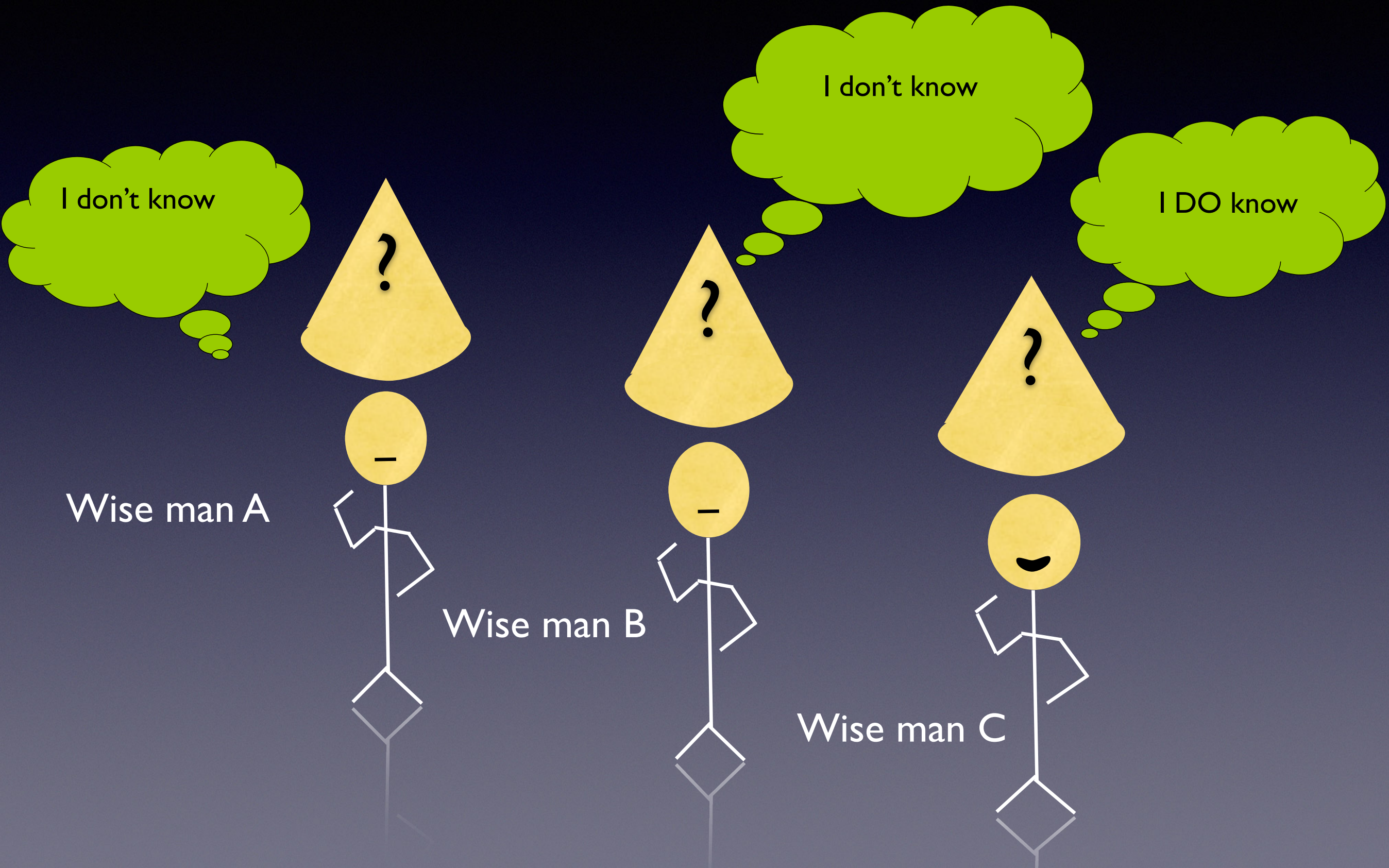


# Wise Men Puzzle



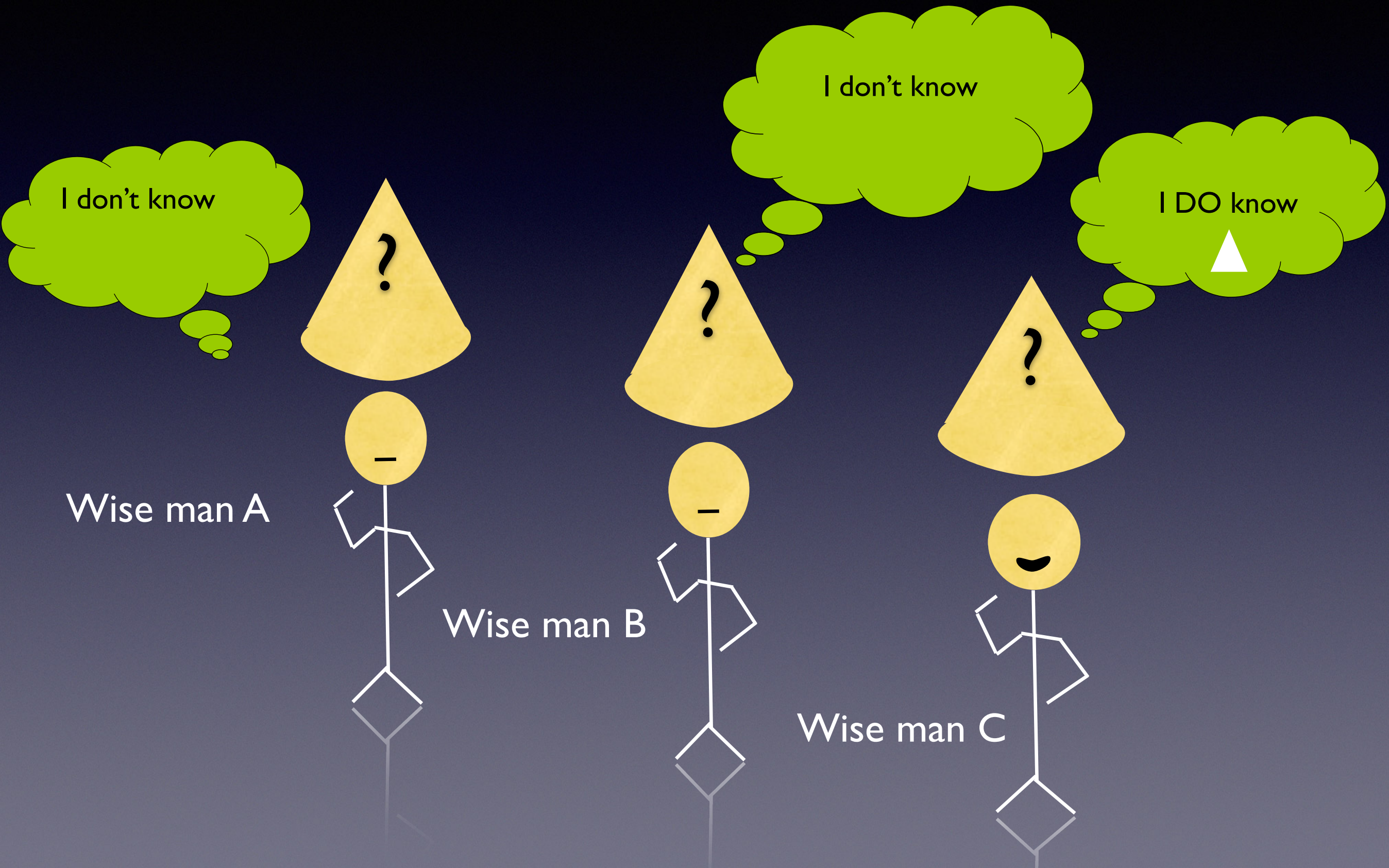


# Wise Men Puzzle



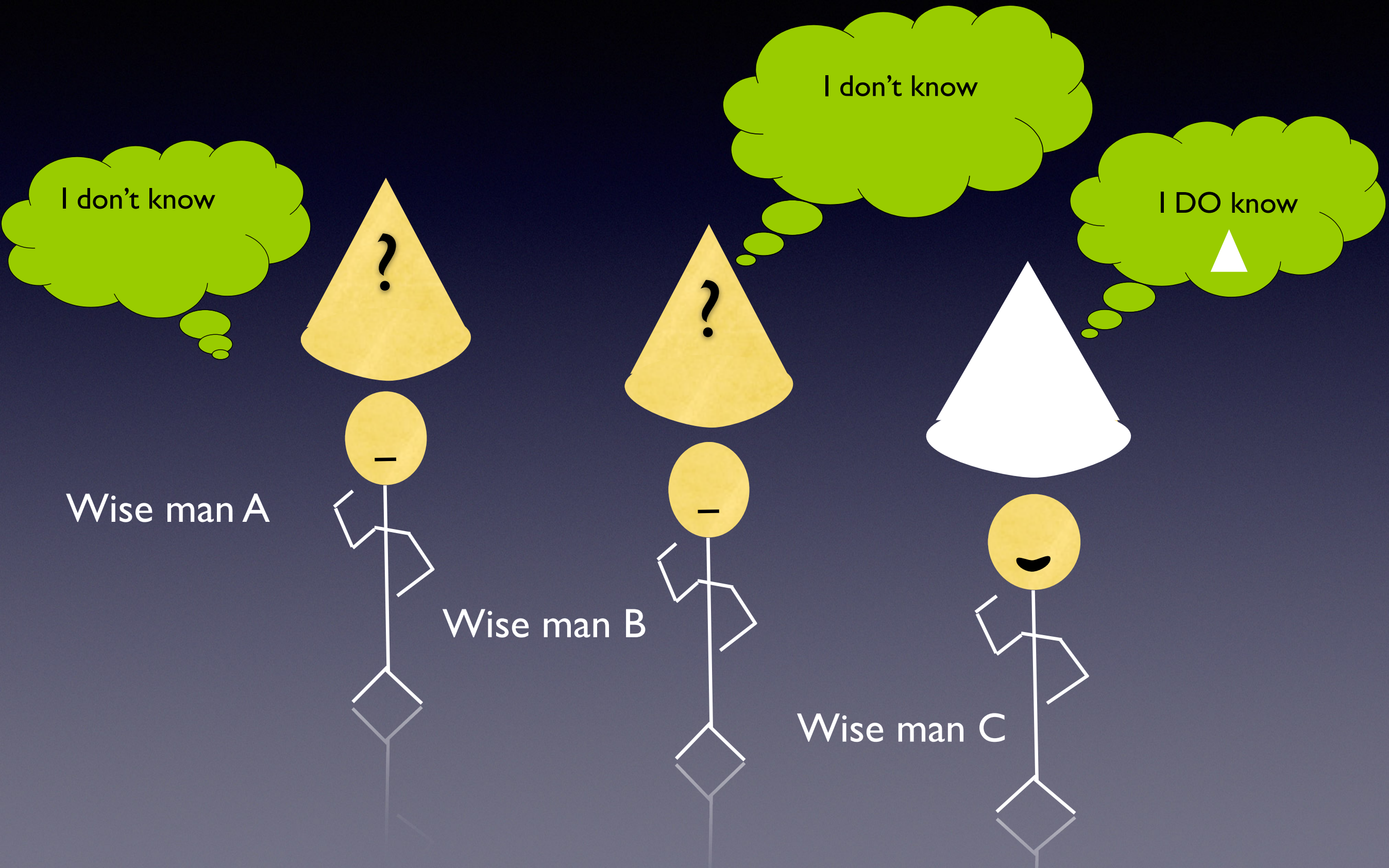


# Wise Men Puzzle





# Wise Men Puzzle





# Proof from WM3

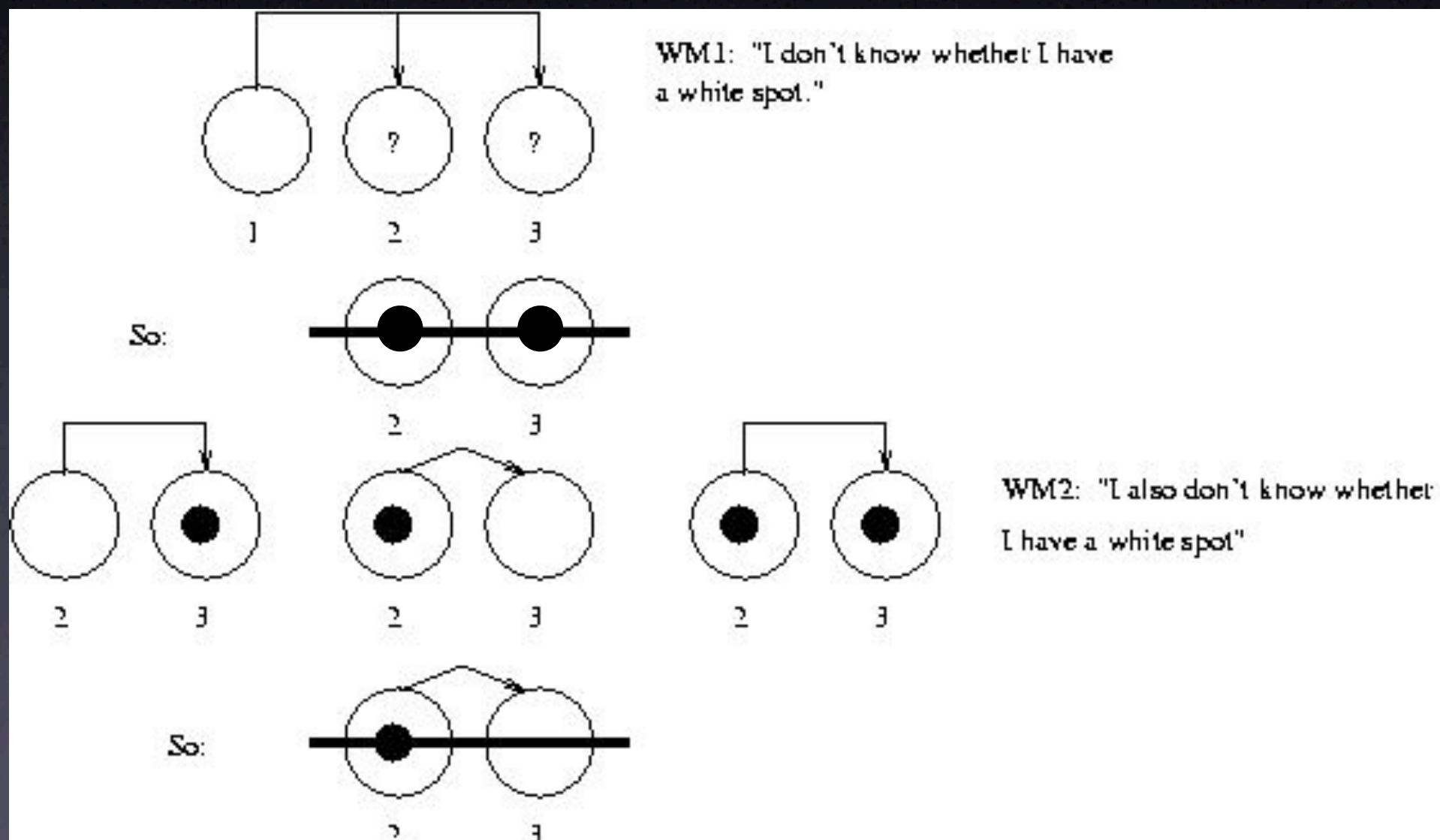
**Proposition:** I have a white fez.

**Proof:** Remember as a first fact that we all know that at least one of us has a white fez. When the first wise man says that he doesn't know, I immediately know that either WM2 has a white fez, or I do, or both of us does. I know this because if neither of us has a white fez, WM1 would have said immediately that in light of our first fact, he has a white fez. My next piece of info comes from what WM2 says; he says that he is *also* ignorant. Now, if he had seen no white fez on my head, he would have immediately said "I have a white fez!" (He would have said this because after WM1 spoke, he carried out the same reasoning I did, and hence ruled out the (WM2-bf & WM3-bf) permutation.) But this *isn't* what he said. Hence, I do have a white fez on my head. QED



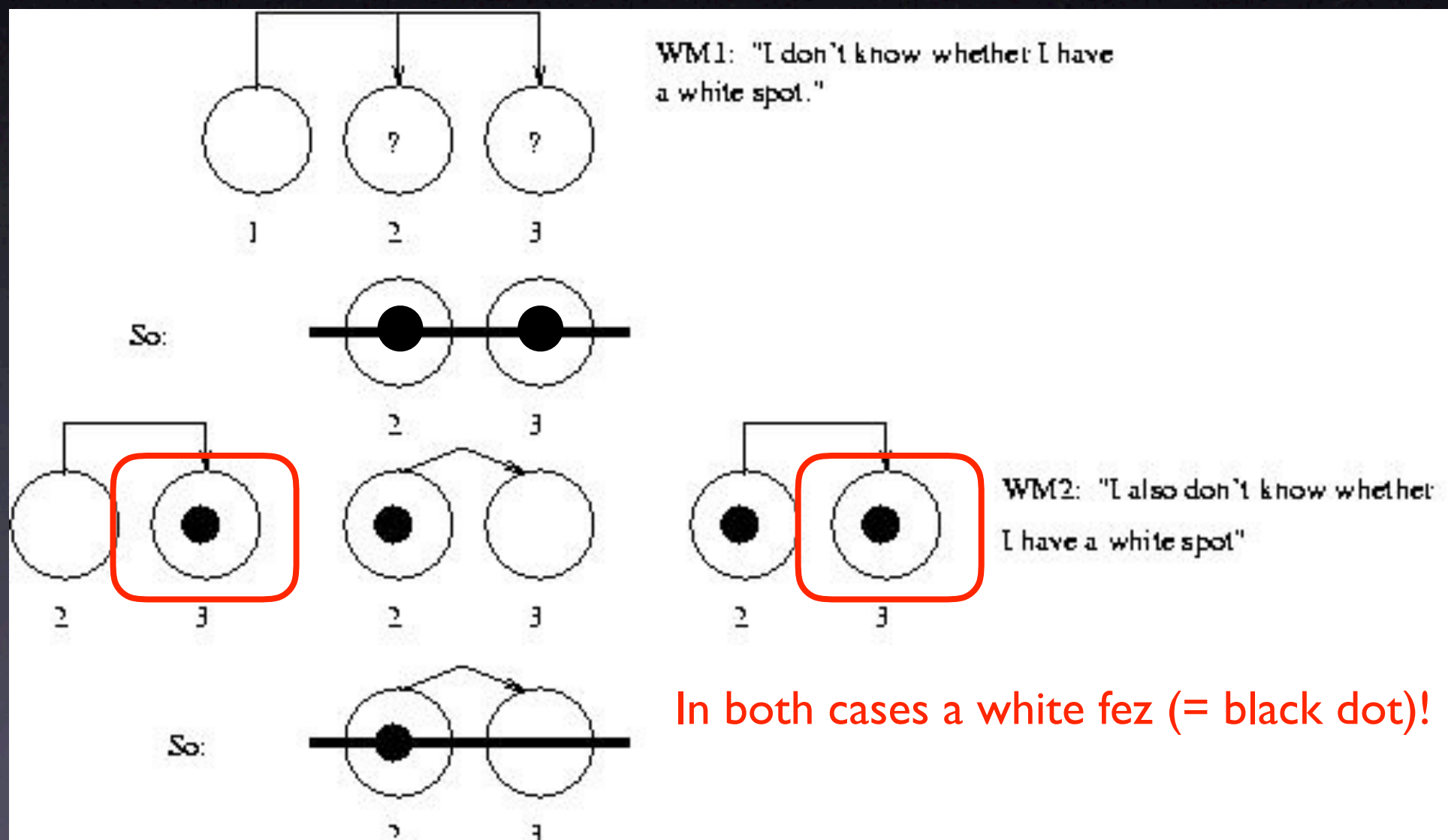
# Diagrammatic Version of Reasoning in WMP<sub>3</sub>

(pov of *truly* wise man; easy for rational humans)



# Diagrammatic Version of Reasoning in WMP<sub>3</sub>

(pov of *truly* wise man; easy for rational humans)



**Abstract.** We present an encoding of a sequent calculus for a multi-agent epistemic logic in Athena, an interactive theorem proving system for many-sorted first-order logic. We then use Athena as a metalanguage in order to reason about the multi-agent logic as an object language. This facilitates theorem proving in the multi-agent logic in several ways. First, it lets us marshal the highly efficient theorem provers for classical first-order logic that are integrated with Athena for the purpose of doing proofs in the multi-agent logic. Second, unlike model-theoretic embeddings of modal logics into classical first-order logic, our proofs are directly convertible into native epistemic logic proofs. Third, because we are able to quantify over propositions and agents, we get much of the generality and power of higher-order logic even though we are in a first-order setting. Finally, we are able to use Athena's versatile tactics for proof automation in the multi-agent logic. We illustrate by developing a tactic for solving the generalized version of the wise men problem.

## 1 Introduction

Multi-agent modal logics are widely used in Computer Science and AI. Multi-agent epistemic logics, in particular, have found applications in fields ranging from AI domains such as robotics, planning, and motivation analysis in natural language [13]; to negotiation and game theory in economics; to distributed systems analysis and protocol authentication in computer security [16, 31]. The reason is simple—intelligent agents must be able to reason about knowledge. It is therefore important to have efficient means for performing machine reasoning in such logics. While the validity problem for most propositional modal logics is of intractable theoretical complexity<sup>1</sup>, several approaches have been investigated in recent years that have resulted in systems that appear to work well in practice. These approaches include tableau-based provers, SAT-based algorithms, and translations to first-order logic coupled with the use of resolution-based automated theorem provers (ATPs). Some representative systems are FuCT [24], KSATC [14], TA [25], LWB [23], and MSPASS [37].

Translation-based approaches (such as that of MSPASS) have the advantage of leveraging the tremendous implementation progress that has occurred over

<sup>1</sup> For instance, the validity problem for multi-agent propositional epistemic logic is PSPACE-complete [18]; adding a common knowledge operator makes the problem EXPTIME-complete [21].

# Arkoudas-Proved-Sound Algorithm for Generating Proof-Theoretic Solution to WMP<sub>n</sub>

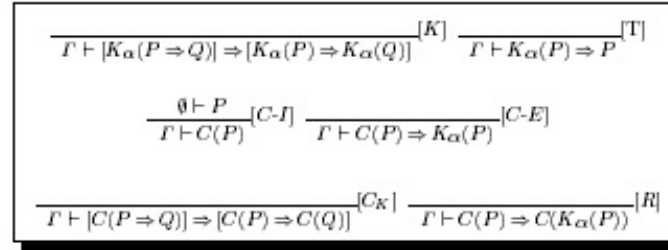


Fig. 2. Inference rules for the epistemic operators.

is  $\Gamma \vdash P$ . Intuitively, this is a judgment stating that  $P$  follows from  $\Gamma$ . We will write  $P, \Gamma$  (or  $\Gamma, P$ ) as an abbreviation for  $\Gamma \cup \{P\}$ . The sequent calculus that we will use consists of a collection of inference rules for deriving judgments of the form  $\Gamma \vdash P$ . Figure 1 shows the inference rules that deal with the standard propositional connectives. This part is standard (e.g., it is very similar to the sequent calculus of Ebbinghaus et al. [15]). In addition, we have some rules pertaining to  $K_\alpha$  and  $C$ , shown in Figure 2.

Rule  $[K]$  is the sequent formulation of the well-known *Kripke axiom* stating that the knowledge operator distributes over conditionals. Rule  $[C_K]$  is the corresponding principle for the common knowledge operator. Rule  $[T]$  is the “truth axiom”: an agent cannot know false propositions. Rule  $[C_I]$  is an introduction rule for common knowledge: if a proposition  $P$  follows from the empty set of hypotheses, i.e., if it is a tautology, then it is commonly known. This is the common-knowledge version of the “omniscience axiom” for single-agent knowledge which says that  $\Gamma \vdash K_\alpha(P)$  can be derived from  $\emptyset \vdash P$ . We do not need to postulate that axiom in our formulation, since it follows from  $[C-I]$  and  $[C-E]$ . The latter says that if it is common knowledge that  $P$  then any (every) agent knows  $P$ , while  $[R]$  says that if it is common knowledge that  $P$  then it is common knowledge that (any) agent  $\alpha$  knows it.  $[R]$  is a reiteration rule that allows us to capture the recursive behavior of  $C$ , which is usually expressed via the so-called “induction axiom”

$$C(P \Rightarrow E(P)) \Rightarrow [P \Rightarrow C(P)]$$

where  $E$  is the shared-knowledge operator. Since we do not need  $E$  for our purposes, we omit its formalization and “unfold”  $C$  via rule  $[R]$  instead.

We state a few lemmas that will come handy later:

**Lemma 1 (Cut).** *If  $\Gamma_1 \vdash P_1$  and  $\Gamma_2, P_1 \vdash P_2$  then  $\Gamma_1 \cup \Gamma_2 \vdash P_2$ .*

**Proof:** Assume  $\Gamma_1 \vdash P_1$  and  $\Gamma_2, P_1 \vdash P_2$ . Then, by  $[ \Rightarrow -I ]$ , we get  $\Gamma_2 \vdash P_1 \Rightarrow P_2$ . Further, by dilation, we have  $\Gamma_1 \cup \Gamma_2 \vdash P_1 \Rightarrow P_2$  and  $\Gamma_1 \cup \Gamma_2 \vdash P_1$ . Hence, by  $[ \Rightarrow -E ]$ , we obtain  $\Gamma_1 \cup \Gamma_2 \vdash P_2$ .  $\square$

The proofs of the remaining lemmas are equally simple exercises:

$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash R_1$	$[Reflex], \wedge-E_1$
$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash R_2$	$[Reflex], \wedge-E_1, \wedge-E_2$
$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash R_3$	$[Reflex], \wedge-E_2$
$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash K_\alpha(\neg Q) \Rightarrow K_\alpha(P)$	2, $[K], \Rightarrow-E$
$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash \neg Q \Rightarrow K_\alpha(P)$	3, 4, Lemma 2
$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash \neg K_\alpha(P) \Rightarrow \neg \neg Q$	5, Lemma 3
$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash \neg \neg Q$	6, 1, $\Rightarrow-E$
$\mathcal{V}_1 \wedge R_2 \wedge R_3 \vdash Q$	7, $[ \neg -E ]$

$\square$

at the above proof is not entirely low-level because most steps combine more inference rule applications in the interest of brevity.

**a 7.** Consider any agent  $\alpha$  and propositions  $P, Q$ . Define  $R_1$  and  $R_3$  as in Lemma 6, let  $R_2 = P \vee Q$ , and let  $S_i = C(R_i)$  for  $i = 1, 2, 3$ . Then  $S_3 \vdash C(Q)$ .

Let  $R'_2 = \neg Q \Rightarrow P$  and consider the following derivation:

$S_1, S_2, S_3 \vdash S_1$	$[Reflex]$
$S_1, S_2, S_3 \vdash S_2$	$[Reflex]$
$S_1, S_2, S_3 \vdash S_3$	$[Reflex]$
$\vdash (P \vee Q) \Rightarrow (\neg Q \Rightarrow P)$	Lemma 4a
$S_1, S_2, S_3 \vdash C((P \vee Q) \Rightarrow (\neg Q \Rightarrow P))$	4, $[C-I]$
$S_1, S_2, S_3 \vdash C(P \vee Q) \Rightarrow C(\neg Q \Rightarrow P)$	5, $[C_K], [ \Rightarrow -E ]$
$S_1, S_2, S_3 \vdash C(\neg Q \Rightarrow P)$	6, 2, $[ \Rightarrow -E ]$
$S_1, S_2, S_3 \vdash C(\neg Q \Rightarrow P) \Rightarrow C(K_\alpha(\neg Q \Rightarrow P))$	$[R]$
$S_1, S_2, S_3 \vdash C(K_\alpha(\neg Q \Rightarrow P))$	8, 7, $[ \Rightarrow -E ]$
$R_1 \wedge K_\alpha(\neg Q \Rightarrow P) \wedge R_3 \vdash Q$	Lemma 6
$\vdash (R_1 \wedge K_\alpha(\neg Q \Rightarrow P) \wedge R_3) \Rightarrow Q$	10, $[ \Rightarrow -I ]$
$S_1, S_2, S_3 \vdash C((R_1 \wedge K_\alpha(\neg Q \Rightarrow P) \wedge R_3) \Rightarrow Q)$	11, $[C-I]$
$S_1, S_2, S_3 \vdash C(R_1 \wedge K_\alpha(\neg Q \Rightarrow P) \wedge R_3) \Rightarrow C(Q)$	12, $[C_K], [ \Rightarrow -E ]$
$S_1, S_2, S_3 \vdash C(R_1 \wedge K_\alpha(\neg Q \Rightarrow P) \wedge R_3)$	1, 3, 9, Lemma 5, $[ \wedge -I ]$
$S_1, S_2, S_3 \vdash C(Q)$	13, 14, $[ \Rightarrow -E ]$

$\square$

all  $n \geq 1$ , it turns out that the last— $(n + 1)^{st}$ —wise man knows he is . The case of two wise men is simple. The reasoning runs essentially by induction. The second wise man reasons as follows:

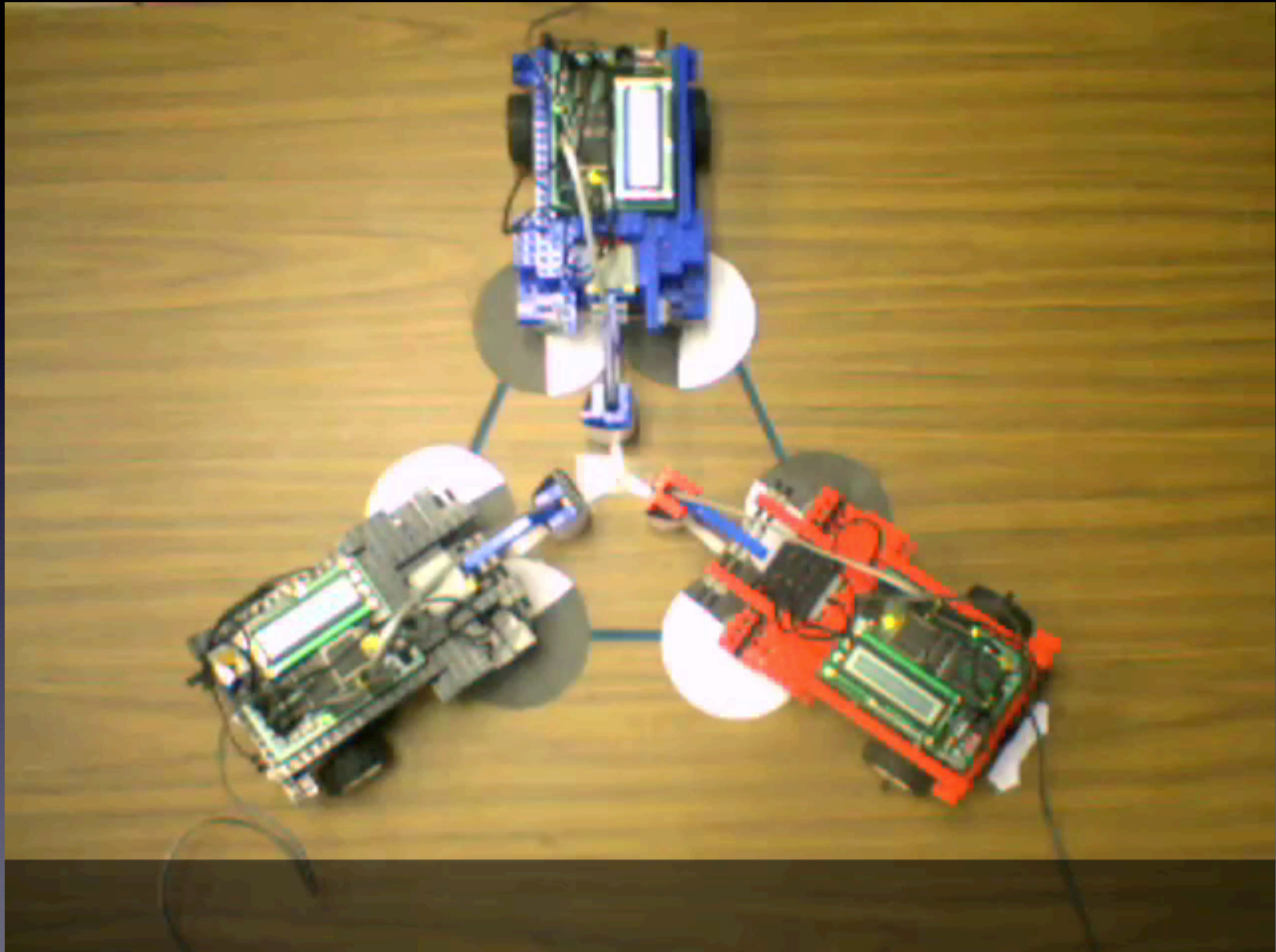
pose I were not marked. Then  $w_1$  would have seen this, and knowing that at least one of us is marked, he would have inferred that he was marked one. But  $w_1$  has expressed ignorance; therefore, I must be marked.

For now the case of  $n = 3$  wise men  $w_1, w_2, w_3$ . After  $w_1$  announces that he does not know that he is marked,  $w_2$  and  $w_3$  both infer that at least one of them is marked. For if neither  $w_2$  nor  $w_3$  were marked,  $w_1$  would have seen this and would have concluded—and stated—that he was the marked one, since he knows that at least one of the three is marked. At this point the puzzle reduces to the two-men case: both  $w_2$  and  $w_3$  know that at least one of them is marked,

All our  
human-  
authored  
proofs  
machine-  
checked.



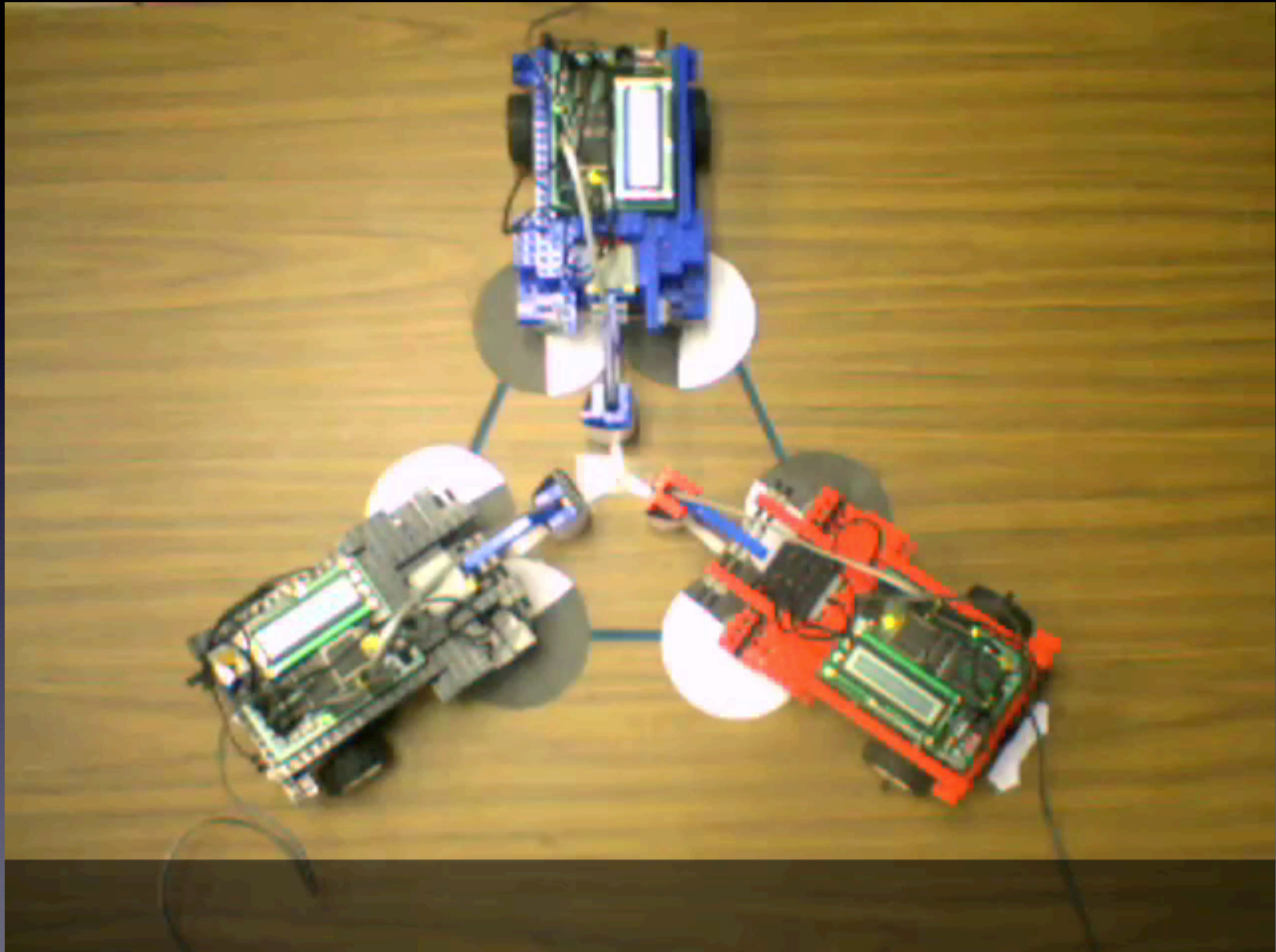
# “Life and Death” Wise Man Test (3)



\* Again: Object-level reasoning, reasoning that *produces* object-level reasoning (e.g., methods), and direct, “dirty,” purely computational procedures.



# “Life and Death” Wise Man Test (3)



\* Again: Object-level reasoning, reasoning that *produces* object-level reasoning (e.g., methods), and direct, “dirty,” purely computational procedures.

Now, harder ...



# Floridi's Continuum (augmented), and Claims

	False Belief Task	Wise Man Test ( <i>n</i> )	Deafening Test	Torture Boots Test	Ultimate Sifter	<i>Infinitary</i> False Belief Task
Cutting-Edge AI	Yes	Yes	No	No	No	?
Zombies	Yes	Yes	Yes	Yes	No	?
Human Persons (s-conscious! p-conscious!)	Yes	Yes	Yes	Yes	Yes	Yes

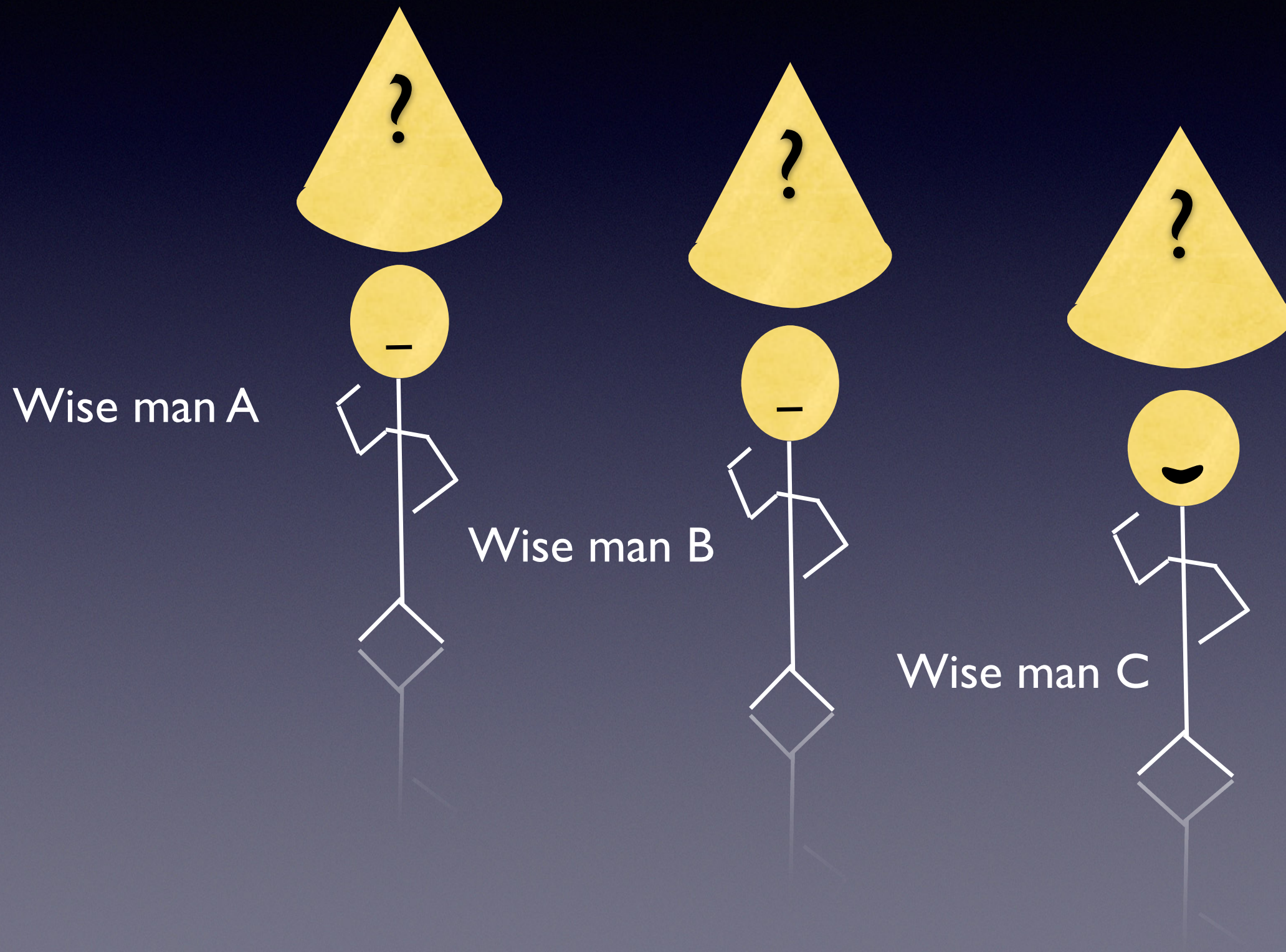


# Floridi's Continuum (augmented), and Claims

	False Belief Task	Wise Man Test ( <i>n</i> )	Deafening Test	Torture Boots Test	Ultimate Sifter	<i>Infinitary</i> False Belief Task
Cutting-Edge AI	Yes	Yes	No	No	No	?
Zombies	Yes	Yes	Yes	Yes	No	?
Human Persons (s-conscious! p-conscious!)	Yes	Yes	Yes	Yes	Yes	Yes



# Floridi's “Ultimate (s- and p-consciousness) Sifter”

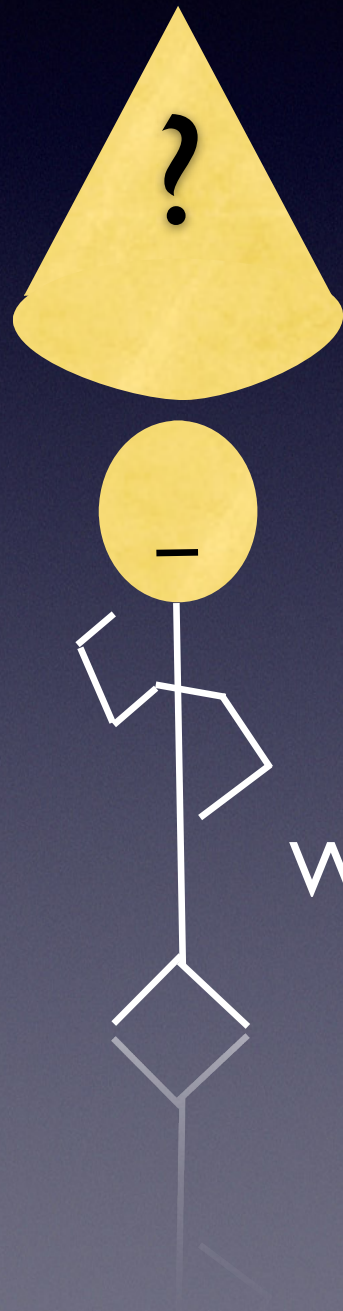




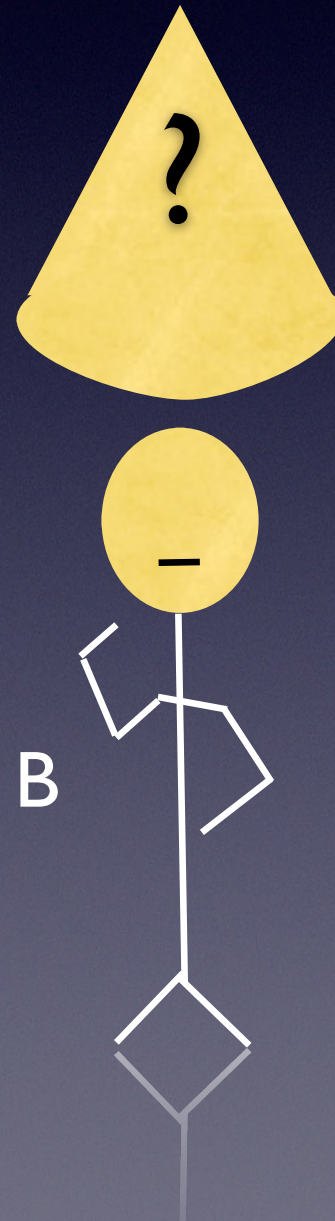
poison innocuous



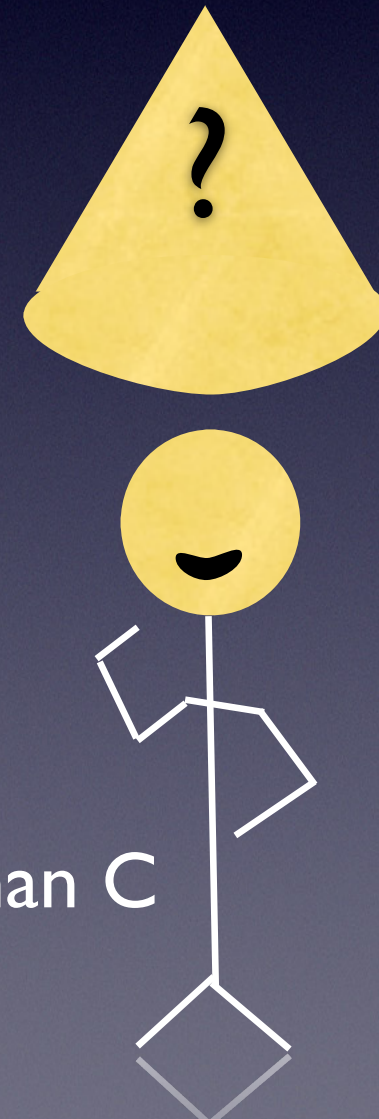
Wise man A



Wise man B



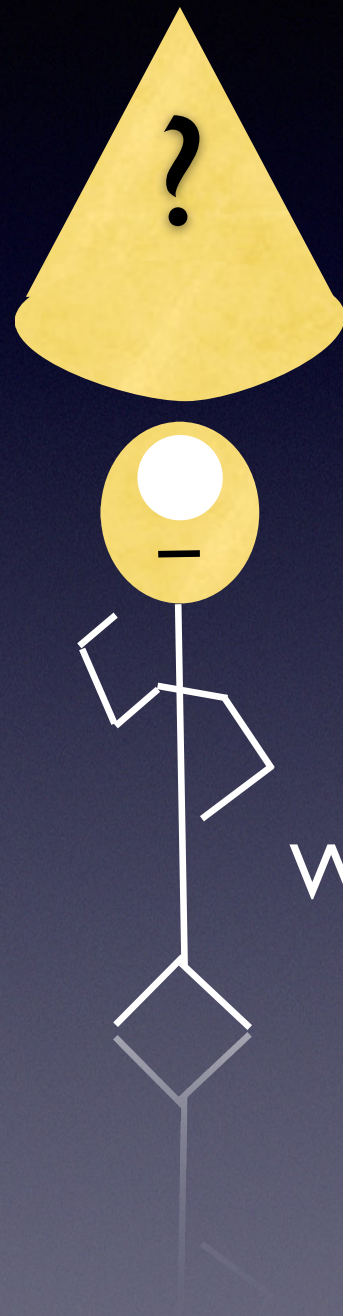
Wise man C



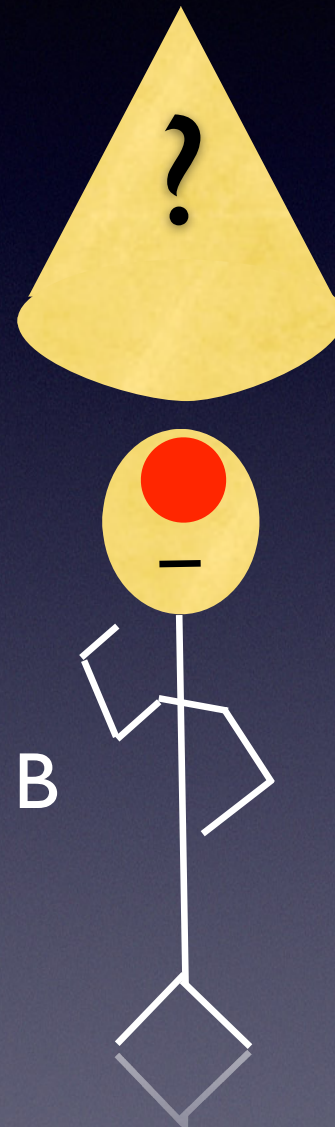
Poison pill strikes the taker dumb.



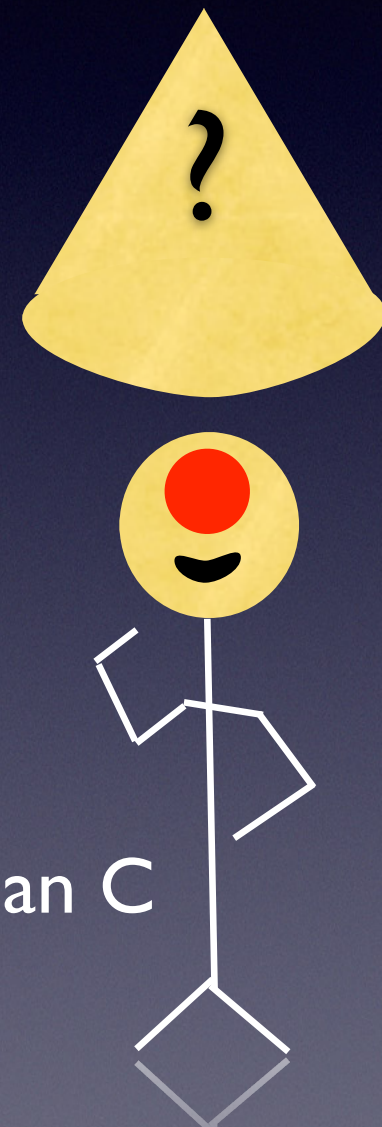
Wise man A



Wise man B

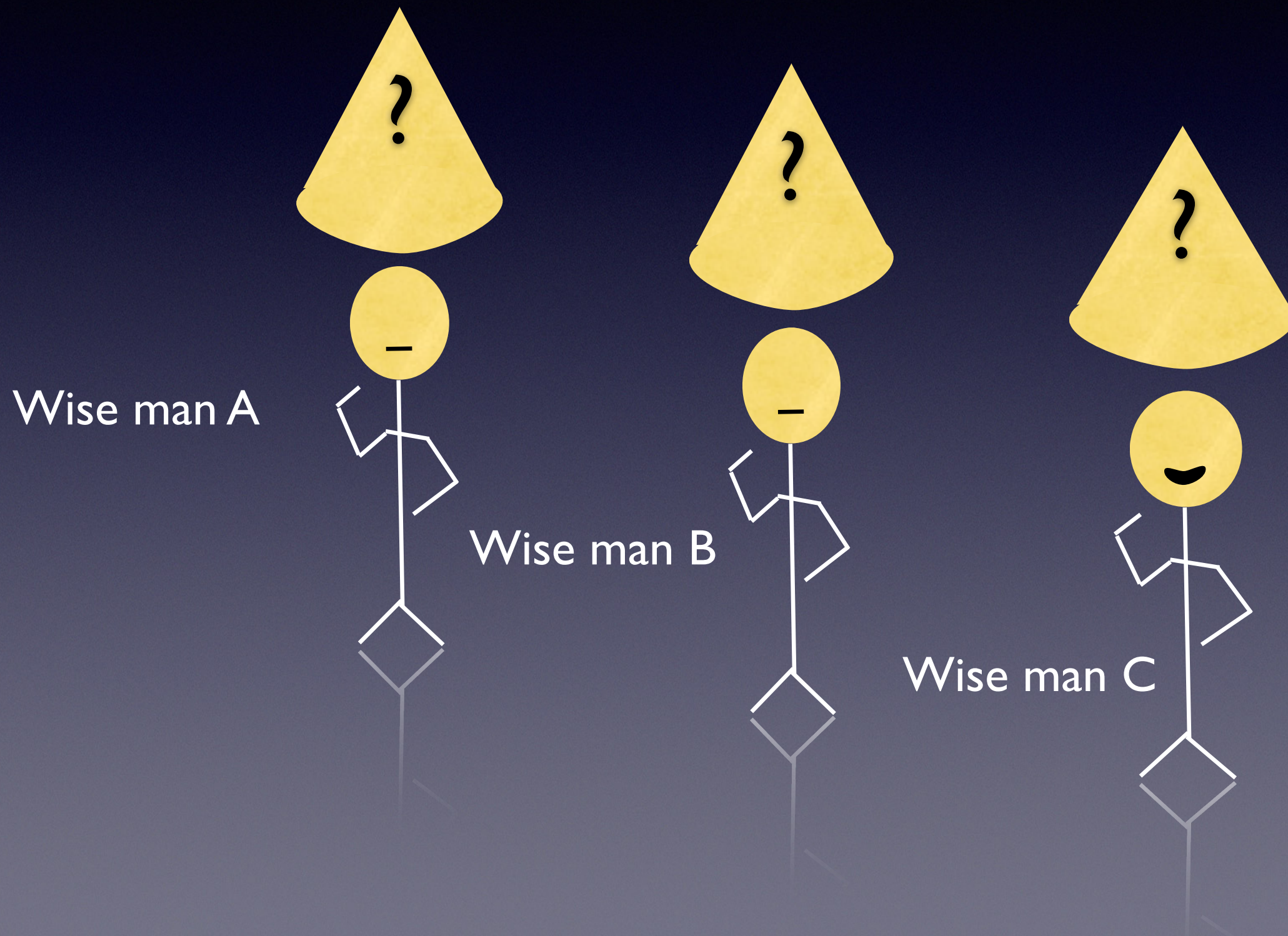


Wise man C





“Have you been struck dumb?  
As always: Prove it!”

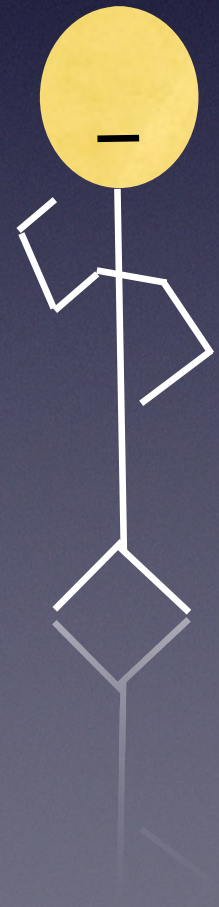




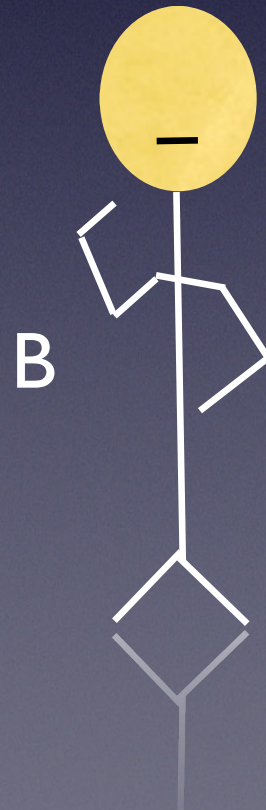
# “Have you been struck dumb?”

Heaven knows!

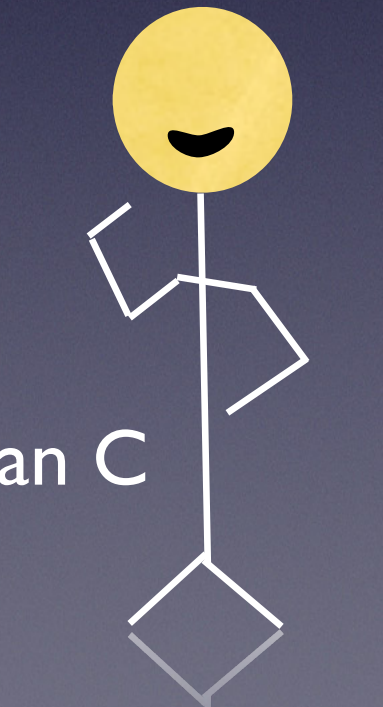
Wise man A



Wise man B



Wise man C





Two possibilities:

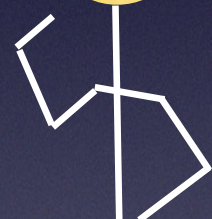
Subsequent silence: failure/death.

Or ...





Wise man A



Wise man B



Wise man C



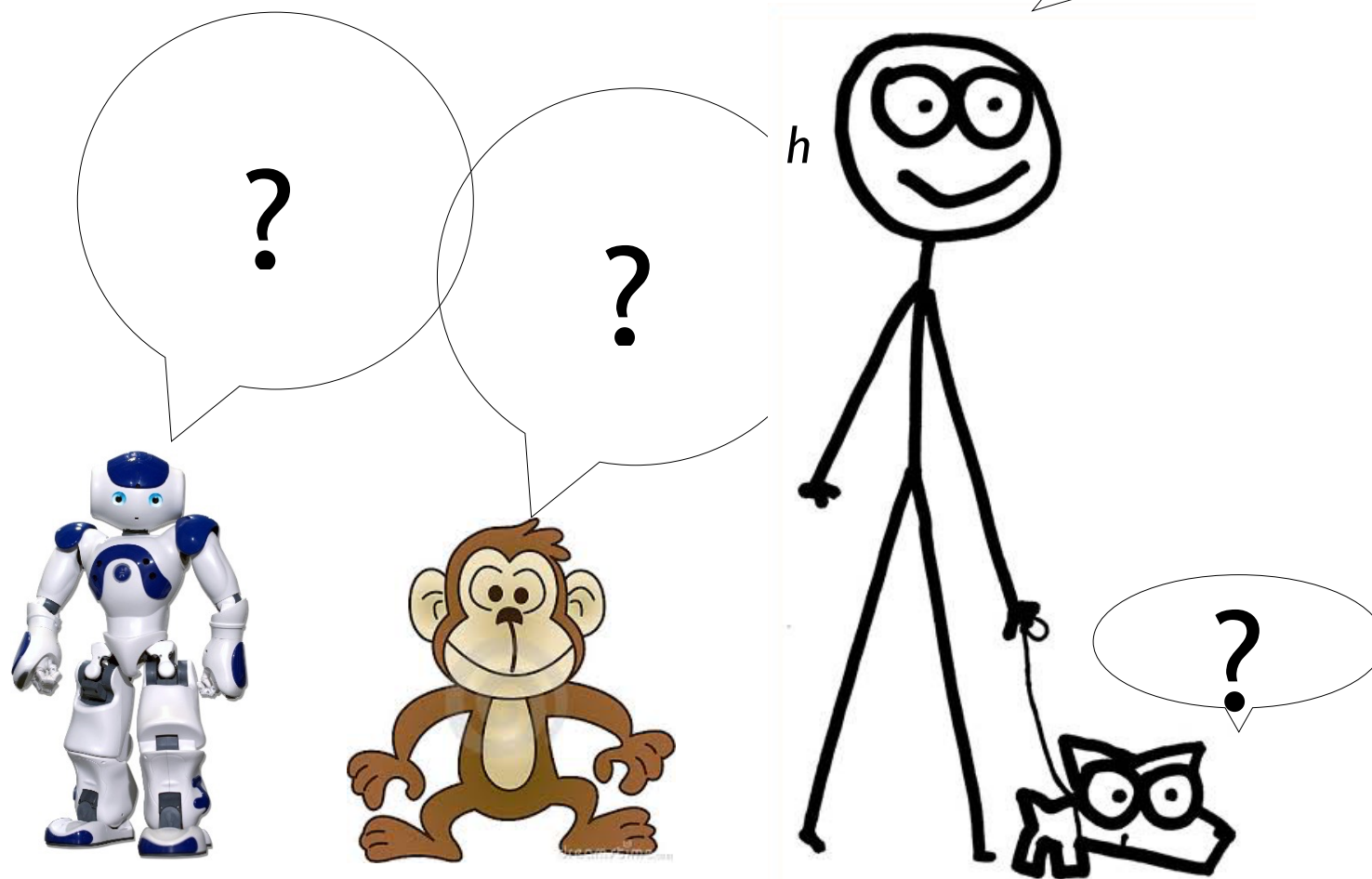


“Had I taken the dumbing tablet I would not have been able to report orally my state of ignorance about my dumb/non-dumb state, but I have been, and I know that I have been, as I have heard myself speaking and saw the guard reacting to my speaking, but this (my oral report) is possible only if I did not take the dumbing tablet, so I know I know I am in the non-dumb state, hence I know that ...”

—Luciano Floridi



$$\langle \chi, \pi \rangle \rightsquigarrow \langle \alpha, \textit{argument/proof} \rangle$$

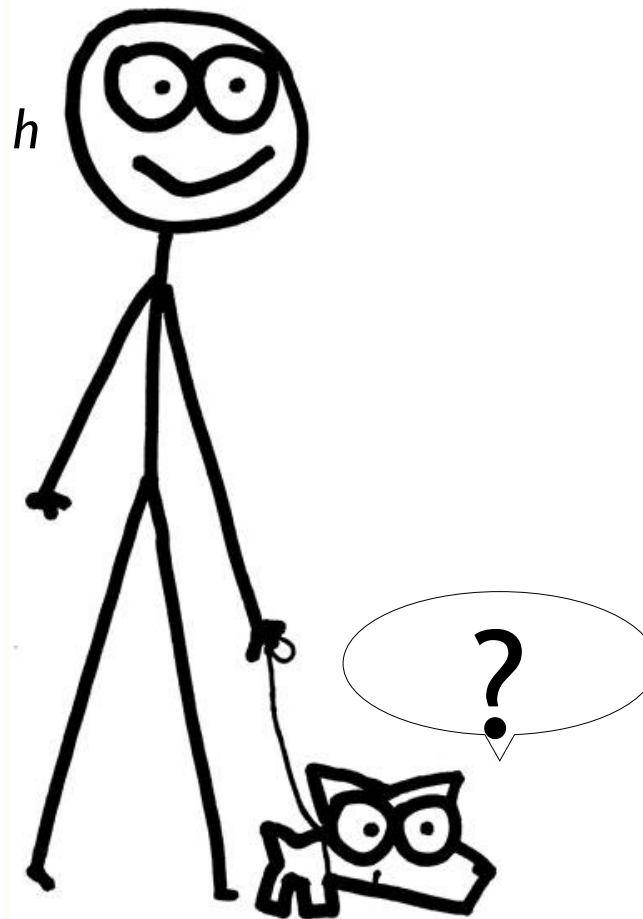


Contrarian view on animal minds in *Nat. Geo.*:

<http://ngm.nationalgeographic.com/2008/03/animal-minds/virginia-morell-text>



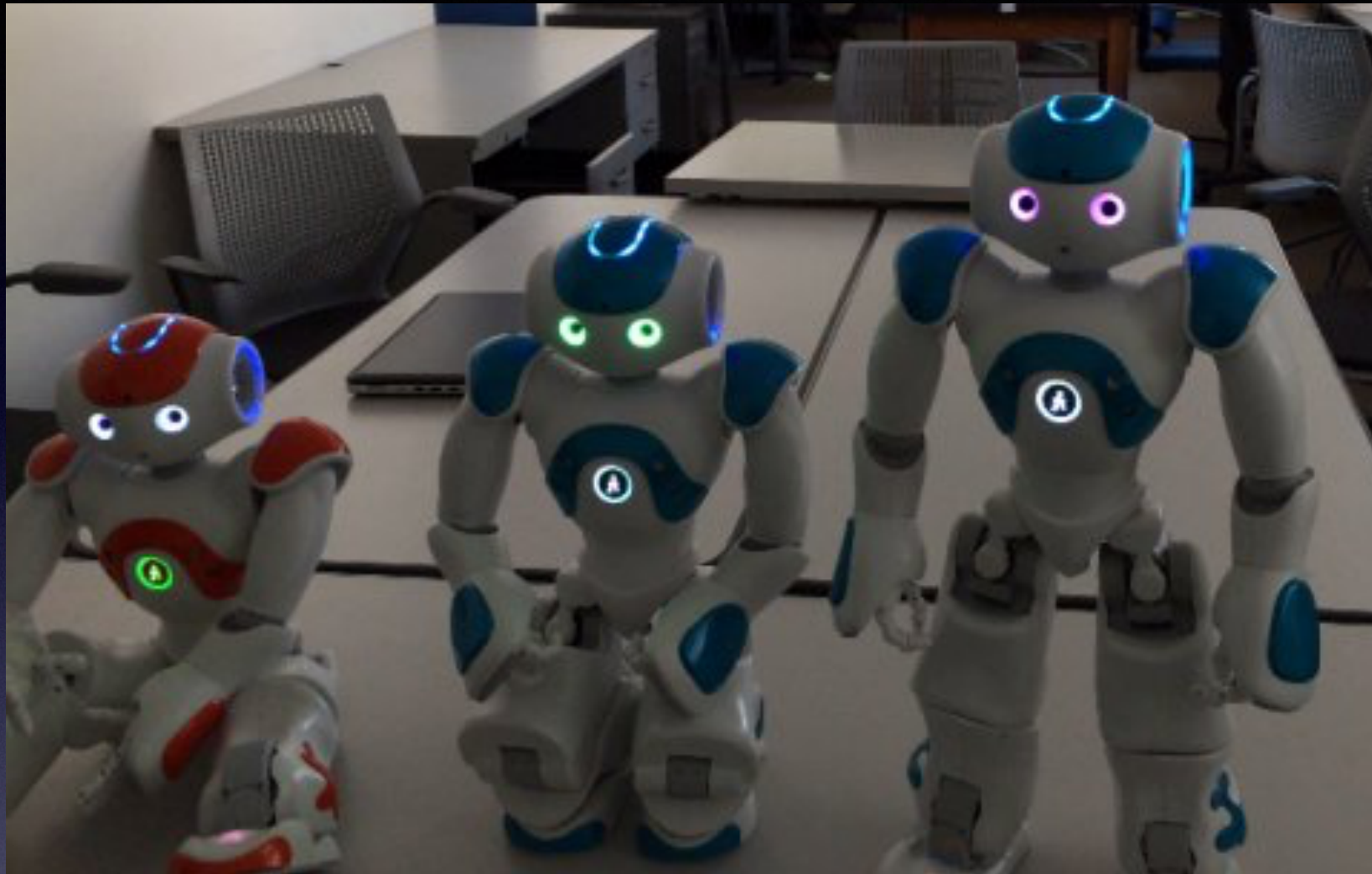
$$\langle \chi, \pi \rangle \rightsquigarrow \langle \alpha, \textit{argument/proof} \rangle$$



Contrarian view on animal minds in *Nat. Geo.*:

<http://ngm.nationalgeographic.com/2008/03/animal-minds/virginia-morell-text>





[http://kryten.mm.rpi.edu/SBringsjord\\_etal\\_self-con\\_robots\\_kg4\\_0601151615NY.pdf](http://kryten.mm.rpi.edu/SBringsjord_etal_self-con_robots_kg4_0601151615NY.pdf)

[https://www.youtube.com/watch?v=MceJYhVD\\_xY](https://www.youtube.com/watch?v=MceJYhVD_xY)



# Floridi's Continuum (augmented), and Claims

	False Belief Task	Wise Man Test ( <i>n</i> )	Deafening Test	Torture Boots Test	Ultimate Sifter	<i>Infinitary</i> False Belief Task
Cutting-Edge AI	Yes	Yes	No	No	No	?
Zombies	Yes	Yes	Yes	Yes	No	?
Human Persons (s-conscious! p-conscious!)	Yes	Yes	Yes	Yes	Yes	Yes



# Floridi's Continuum (augmented), and Claims

	False Belief Task	Wise Man Test ( <i>n</i> )	Deafening Test	Torture Boots Test	Ultimate Sifter	<i>Infinitary</i> False Belief Task
Cutting-Edge AI	Yes	Yes	No	No	No	?
Zombies	Yes	Yes	Yes	Yes	No	?
Human Persons (s-conscious! p-conscious!)	Yes	Yes	Yes	Yes	Yes	Yes



# Infinitary False Belief Task

<http://kryten.mm.rpi.edu/PRES/COGSCI2019/infinitaryfalsebeliefprezCogSci2019.key>



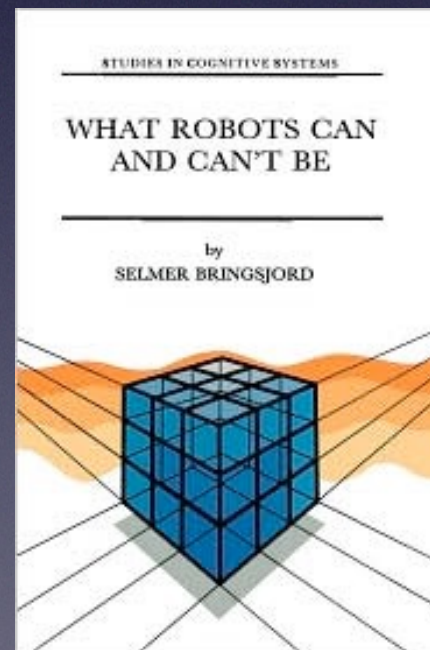




So ... despite the fact we can't build rational persons,  
apparently we can build AIs that pass *any* short test.  
That's why *Blade Runner* (& *Ex Machina*?) is our future.



So ... despite the fact we can't build rational persons,  
apparently we can build AIs that pass *any* short test.  
That's why *Blade Runner* (& *Ex Machina*?) is our future.





# Refutation of: Cognitive Attack $\mathcal{R}$

If humans are as described in this thesis, then they can solve the forthcoming cognitive problems.

But humans *can't* solve the problems in question.

Therefore:

Sorry Selmer & company, your thesis  $\mathcal{R}$  is false.



# Refutation of: Cognitive Attack $\mathcal{R}$

If humans are as described in this thesis, then they **FALSE** can solve the forthcoming cognitive problems.

But humans *can't* solve the problems in question.

Therefore:

Sorry Selmer & company, your thesis  $\mathcal{R}$  is false.



# Refutation of: Cognitive Attack $\mathcal{R}$

If humans are as described in this thesis, then they **FALSE** can solve the forthcoming cognitive problems.

If humans are as described in  $R$ , then humans can, given sufficient training, etc., eventually solve the cognitive problems in question.

But humans *can't* solve the problems in question.

Therefore:

Sorry Selmer & company, your thesis  $\mathcal{R}$  is false.



# Refutation of: Cognitive Attack $\mathcal{R}$

If humans are as described in this thesis, then they **FALSE**  
can solve the forthcoming cognitive problems.

If humans are as described in  $R$ , then humans  
*can, given sufficient training, etc., eventually*  
*solve the cognitive problems in question.*

But humans *can't* solve the problems in question. **FALSE**

Therefore:

Sorry Selmer & company, your thesis  $\mathcal{R}$  is false.



# Refutation of: Cognitive Attack $\mathcal{R}$

If humans are as described in this thesis, then they **FALSE** can solve the forthcoming cognitive problems.

If humans are as described in  $R$ , then humans can, given sufficient training, etc., eventually solve the cognitive problems in question.

But humans *can't* solve the problems in question. **FALSE**

Some humans can't, *at present*, solve the problems in question — & as it turns out, AIs can do surprisingly well, at least until we get to the infinite case.

Therefore:

Sorry Selmer & company, your thesis  $\mathcal{R}$  is false.