Modal Logic Basics

for the Workshop on Logics for Social Behavior

Indiana University

July 12, 2016

My goal today is to beginning NASSLLI participants enough of the basics to follow the next lectures in this workshop,
and of course to inspire and enable further study of topics related to
logic and the social world.

- Knowledge, the standard modeling, done mostly via examples.
- A group exercise to reinforce the ideas
- The simplest dynamic logic
- Common knowledge
- Probability, knowledge, and cascade behavior

We have a deck with three cards ♥, ♣, and ♦.
We also have three players: B, C, and D.

We deal the cards out, one to each player, face down.

We have a deck with three cards ♥, ♣, and ♦.
We also have three players: B, C, and D.

We deal the cards out, one to each player, face down.
For now, let's suppose that the deal was

$$B : \diamondsuit \qquad C : \clubsuit \qquad D : \heartsuit$$

We want a logical language to talk about this scenario, and so we'll use propositional logic.

The main background that you need for the course is an understanding of the symbols of propositional logic, including

$\neg$	not
$\wedge$	and
$\vee$	or
$\rightarrow$	if ... then
$\leftrightarrow$	if and only if

I will review some of this material today and next time, but this is what you need to know well to get started.

We want a logical language to talk about this scenario, and so we'll use propositional logic.
We take as atomic the following nine sentences

$$B\heartsuit, \quad B\clubsuit, \quad B\diamondsuit,$$
$$C\heartsuit, \quad C\clubsuit, \quad C\diamondsuit,$$
$$D\heartsuit, \quad D\clubsuit, \quad D\diamondsuit$$

Remember that the actual deal was

$$(B : \diamondsuit \qquad C : \clubsuit \qquad D : \heartsuit)$$

Which of the following are true at the real world, intuitively?

$$B\heartsuit \lor C\clubsuit$$
$$\neg B\clubsuit$$
$$B\diamondsuit \to (C\clubsuit \lor C\heartsuit)$$

T means "true" and F means "false".
Symbols: $P \wedge Q$: *P and Q*.
$P \vee Q$: *P or Q*.
$P \rightarrow Q$: *P implies Q*, or *If P, then Q*, or *Q, provided that P*.

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

P	Q	$P \vee Q$
T	T	T
T	F	T
F	T	T
F	F	F

P	Q	$P \rightarrow Q$
T	T	T
T	F	F
F	T	T
F	F	T

P	$\neg P$
T	F
F	T

We said that the original deal was

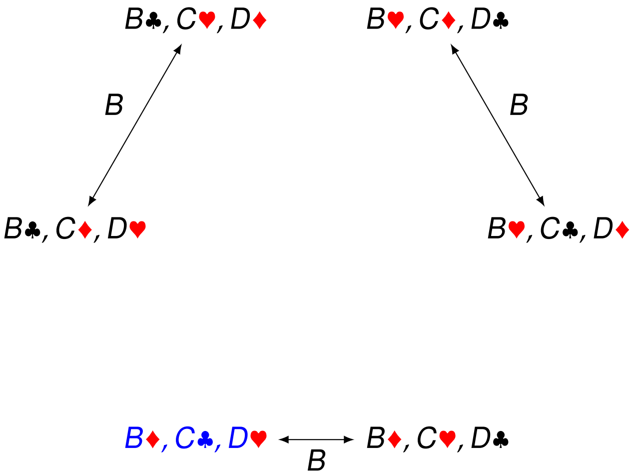
$$(B : \diamondsuit \qquad C : \clubsuit \qquad D : \heartsuit)$$

But this was only one of the six possible deals?
What sentences are true no matter what the original deal was?

$B\clubsuit, C\heartsuit, D\diamondsuit$ $B\heartsuit, C\clubsuit, D\diamondsuit$

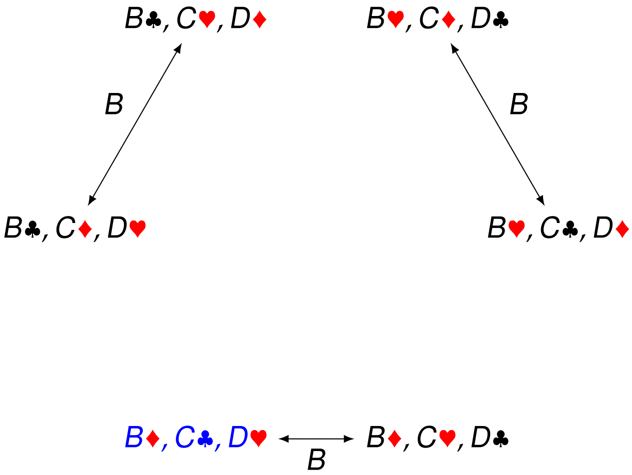
$B\clubsuit, C\diamondsuit, D\heartsuit$ $B\heartsuit, C\diamondsuit, D\clubsuit$

$B\diamondsuit, C\clubsuit, D\heartsuit$ $B\diamondsuit, C\heartsuit, D\clubsuit$


B♣, C♥, D♦
B♥, C♦, D♣
B
B
B♣, C♦, D♥
B♥, C♣, D♦
B♦, C♣, D♥
B♦, C♥, D♣
B

The space of possible deals

with B's indifference relation, and with the real world in blue

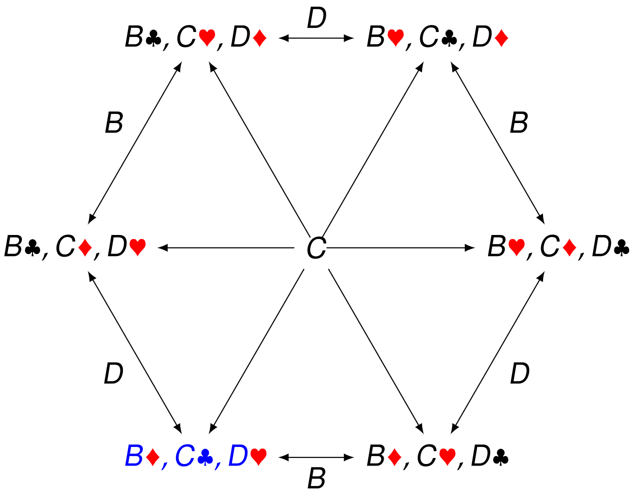


But what do the lines mean?

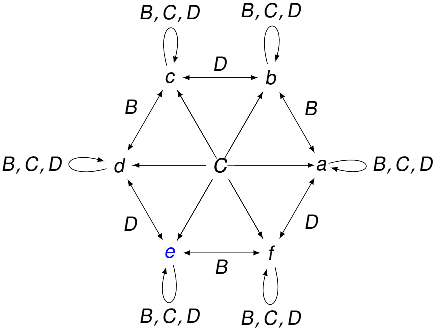
Try to draw the lines for C and for D, all on the same hexagon.

The space of possible deals

with everyone's indifference relation,
but omitting loops on all six nodes, each labeled B, C, and D.
Note also that the arrows through the middle are diagonals.



LET'S WRITE THE WORLDS AS a, b, c, d, e, f. ALSO, THE SELF-LOOPS ARE SHOWN.



Some facts about this model:

$$d \xrightarrow{B} d \quad d \xrightarrow{C} a \quad a \xrightarrow{B} b \quad e \xrightarrow{D} e \quad f \xrightarrow{D} a$$

We would also write $\nrightarrow$ for "not accessible", and then here are some more facts about the model:

$$a \overset{B}{\nrightarrow} d \quad f \overset{C}{\nrightarrow} b \quad a \overset{B}{\nrightarrow} e \quad e \overset{C}{\nrightarrow} c \quad f \overset{D}{\nrightarrow} d$$

Recall that we have a formal language built from atomic sentences

$$B\heartsuit, \quad B\clubsuit, \quad B\diamondsuit,$$
$$C\heartsuit, \quad C\clubsuit, \quad C\diamondsuit,$$
$$D\heartsuit, \quad D\clubsuit, \quad D\diamondsuit$$

using $\wedge$, $\vee$, $\neg$, $\rightarrow$, $\leftrightarrow$, and the knowledge operators K_B, K_C and K_D.

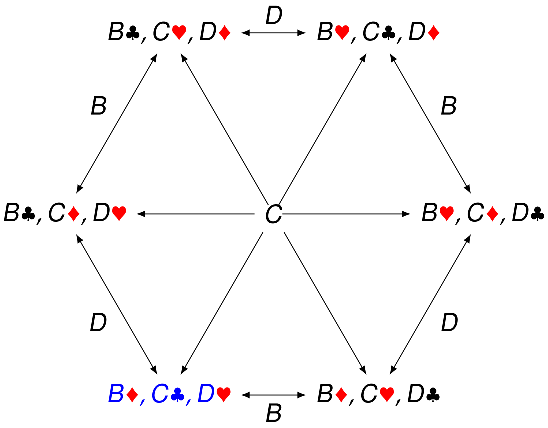
On the next slide, you will see the main formal definition of the semantics of our language.

The points in our pictures are called worlds.
When x is a world in a model:

Read $x \models \varphi$ as x satisfies φ or as φ is true in x

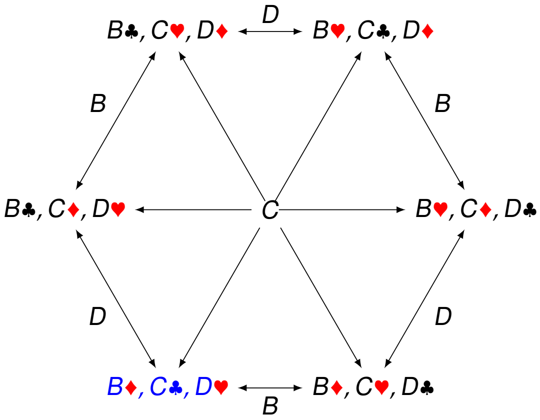
$x \models p$	iff	p is written on x (p atomic)
$x \models \varphi \wedge \psi$	iff	$x \models \varphi$ and $x \models \psi$
$x \models \varphi \vee \psi$	iff	$x \models \varphi$ or $x \models \psi$
$x \models \neg\varphi$	iff	if it is not the case that $x \models \varphi$
$x \models \varphi \rightarrow \psi$	iff	if $x \models \varphi$, then $x \models \psi$
$x \models \varphi \leftrightarrow \psi$	iff	$x \models \varphi$ if and only if $x \models \psi$
$x \models K_B\, \varphi$	iff	$y \models \varphi$ for all y such that $x \xrightarrow{B} y$

The last line is for B, but we mean similar things for C and D, too.



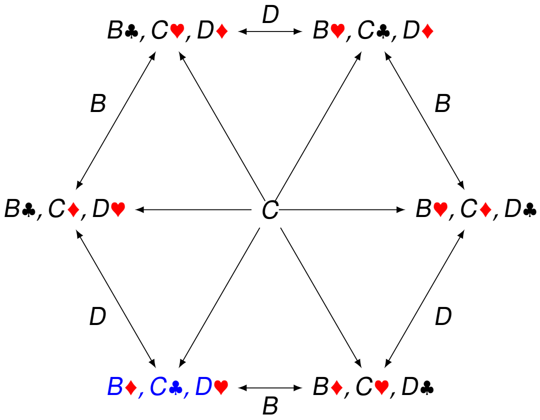
$B\clubsuit, C\diamondsuit, D\heartsuit \models B\clubsuit$	our x here is $B\clubsuit, C\diamondsuit, D\heartsuit$
$B\diamondsuit, C\heartsuit, D\clubsuit \models B\diamondsuit \wedge D\clubsuit$	our x here is $B\diamondsuit, C\heartsuit, D\clubsuit$
$B\clubsuit, C\heartsuit, D\diamondsuit \models \neg(C\clubsuit \vee C\diamondsuit)$	our x here is $B\clubsuit, C\heartsuit, D\diamondsuit$
$B\clubsuit, C\heartsuit, D\diamondsuit \models C\heartsuit \rightarrow D\diamondsuit$	our x here is $B\clubsuit, C\heartsuit, D\diamondsuit$

A harder example



B♦, C♣, D♥ $\models K_B$ B♦

because the only *y* such that B♦, C♣, D♥ is *B*-connected to *y*
are B♦, C♣, D♥ and B♦, C♥, D♣,
and at both of those, B♦ is true.



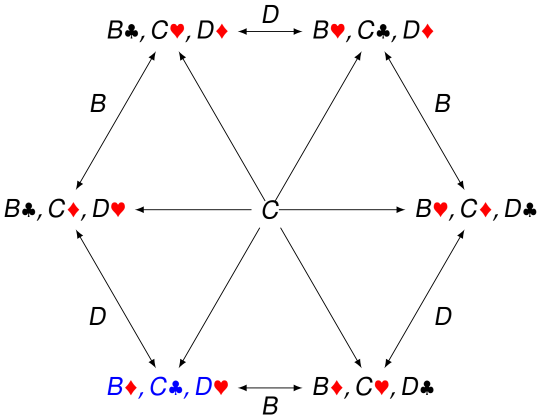
Is it the case that $B\diamondsuit, C\clubsuit, D\heartsuit \models K_B\, C\clubsuit$, or not?
Before we answer this, let's look back at our semantics.

$x \models p$	iff	p is written on x (p atomic)
$x \models \varphi \wedge \psi$	iff	$x \models \varphi$ and $x \models \psi$
$x \models \varphi \vee \psi$	iff	$x \models \varphi$ or $x \models \psi$
$x \models \neg\varphi$	iff	if it is not the case that $x \models \varphi$
$x \models \varphi \rightarrow \psi$	iff	if $x \models \varphi$, then $x \models \psi$
$x \models \varphi \leftrightarrow \psi$	iff	$x \models \varphi$ if and only if $x \models \psi$
$x \models K_B\, \varphi$	iff	$y \models \varphi$ for all y such that $x \xrightarrow{B} y$

$x \not\models p$	iff	p is not written on x (p atomic)
$x \not\models \varphi \wedge \psi$	iff	$x \not\models \varphi$ or $x \not\models \psi$
$x \not\models \varphi \vee \psi$	iff	
$x \not\models \neg\varphi$	iff	
$x \not\models \varphi \rightarrow \psi$	iff	
$x \not\models \varphi \leftrightarrow \psi$	iff	
$x \not\models K_B\, \varphi$	iff	

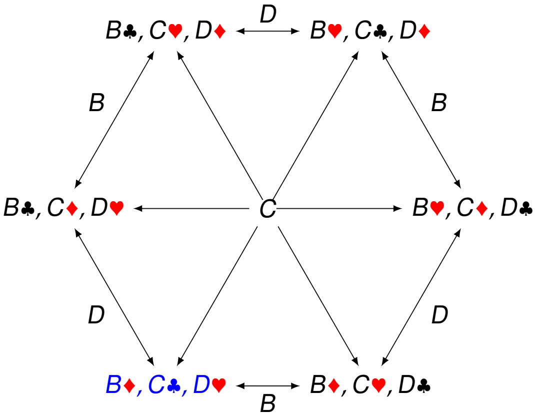
$x \models p$	iff	p is written on x (p atomic)
$x \models \varphi \wedge \psi$	iff	$x \models \varphi$ and $x \models \psi$
$x \models \varphi \vee \psi$	iff	$x \models \varphi$ or $x \models \psi$
$x \models \neg\varphi$	iff	if it is not the case that $x \models \varphi$
$x \models \varphi \rightarrow \psi$	iff	if $x \models \varphi$, then $x \models \psi$
$x \models \varphi \leftrightarrow \psi$	iff	$x \models \varphi$ if and only if $x \models \psi$
$x \models K_B\, \varphi$	iff	$y \models \varphi$ for all y such that $x \xrightarrow{B} y$

$x \not\models p$	iff	p is not written on x (p atomic)
$x \not\models \varphi \wedge \psi$	iff	$x \not\models \varphi$ or $x \not\models \psi$
$x \not\models \varphi \vee \psi$	iff	$x \not\models \varphi$ and $x \not\models \psi$
$x \not\models \neg\varphi$	iff	$x \models \varphi$
$x \not\models \varphi \rightarrow \psi$	iff	$x \models \varphi$, but $x \not\models \psi$
$x \not\models \varphi \leftrightarrow \psi$	iff	($x \models \varphi$, but $x \not\models \psi$) or ($x \models \psi$, but $x \not\models \varphi$)
$x \not\models K_B\, \varphi$	iff	there is some y such that $x \xrightarrow{B} y$, but $y \not\models \varphi$



Is it the case that $B\diamondsuit, C\clubsuit, D\heartsuit \models K_B\, C\clubsuit$, or not?

And does this match our intuitions about what the formal sentence means?



It's not true:
there is some y such that B♦, C♣, D♥ is B-connected to y
(namely B♦, C♥, D♣)
and B♦, C♥, D♣ $\not\models$ C♣.

In the real world (*B*♦, *C*♣, *D*♥), is the following intuitively true?

B knows that if *C* has hearts, then *D* has clubs

In the real world (*B*♦, *C*♣, *D*♥), is the following intuitively true?

B knows that if *C* has hearts, then *D* has clubs

Now formalize this sentence.

In the real world ($B\diamondsuit, C\clubsuit, D\heartsuit$), is the following intuitively true?

B knows that if *C* has hearts, then *D* has clubs

Now formalize this sentence.

Now decide with someone near you if the formal sentence is true at the real world.

In the real world ($B\diamondsuit, C\clubsuit, D\heartsuit$), is the following intuitively true?

B knows that *C* doesn't know that she (*B*) has diamonds

Now formalize this sentence.

Now decide with someone near you if the formal sentence is true at the real world.

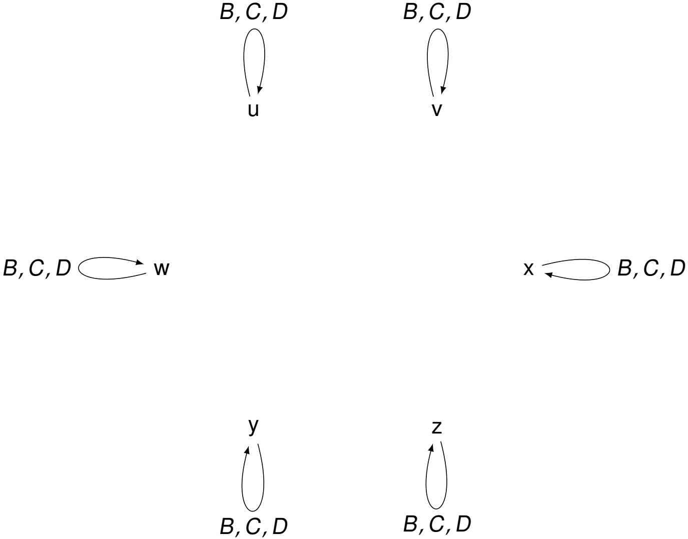
The formalized sentence is $K_B\ \neg K_C\ B\diamondsuit$.
This is true at $B\diamondsuit, C\clubsuit, D\heartsuit$ because:

1. The worlds which B can consider at $B\diamondsuit, C\clubsuit, D\heartsuit$ are $B\diamondsuit, C\clubsuit, D\heartsuit$ and $B\diamondsuit, C\heartsuit, D\clubsuit$.
2. $B\diamondsuit, C\clubsuit, D\heartsuit \not\models K_C\ B\diamondsuit$ because $B\diamondsuit, C\clubsuit, D\heartsuit \overset{C}{—} B\heartsuit, C\clubsuit, D\diamondsuit$ and $B\heartsuit, C\clubsuit, D\diamondsuit \not\models B\diamondsuit$.
3. $B\diamondsuit, C\heartsuit, D\clubsuit \not\models K_C\ B\diamondsuit$ because $B\diamondsuit, C\heartsuit, D\clubsuit \overset{C}{—} B\clubsuit, C\heartsuit, D\diamondsuit$ and $B\clubsuit, C\heartsuit, D\diamondsuit \not\models B\diamondsuit$.
4. By the last two points, $\neg K_C\ B\diamondsuit$ is true in all of the worlds which B can consider at $B\diamondsuit, C\clubsuit, D\heartsuit$.
5. Therefore, at $B\diamondsuit, C\clubsuit, D\heartsuit$, the sentence $K_B\ \neg K_C\ B\diamondsuit$ is true.

We said that the deal was face down.

Let's reconsider this.

Think for a moment about what the space of possible deals would be like if the deal had been face up.

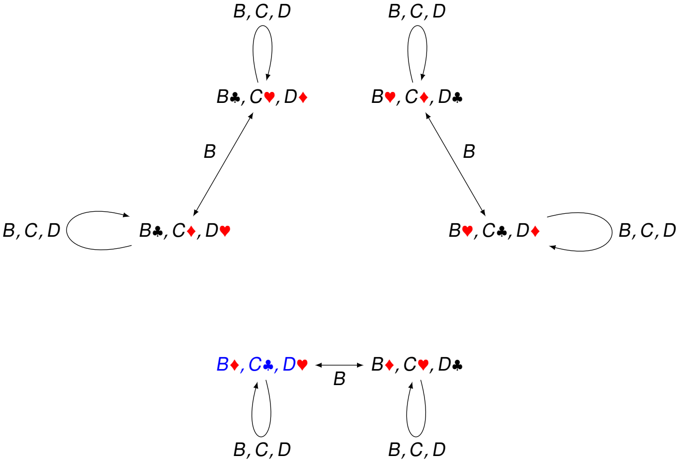

B, C, D
u
B, C, D
v
B, C, D
w
x
B, C, D
y
B, C, D
z
B, C, D

B knows that if *C* has hearts, then *D* has clubs
B knows that *C* doesn't know that she (*B*) has diamonds.

Would they be intuitively true?
Are the formal versions in fact true at the "real world" node?

B just sits there, watching C and D show cards to one another.

What should the representation be?

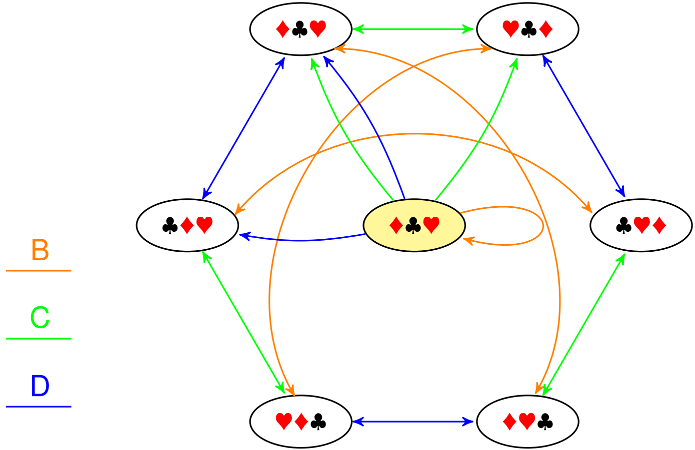


But how can we decide right and wrong answers for this and similar questions?

As the semester goes on, we'll see.

Cheating: B looks at C's card, thereby learning D's.

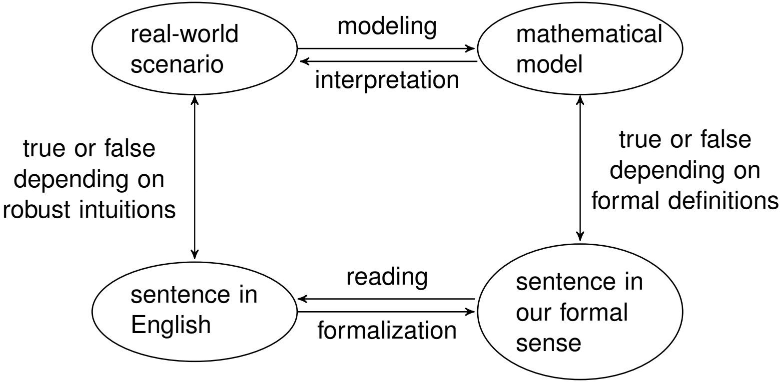
The real world is in the middle.



$B\diamondsuit, C\clubsuit, D\heartsuit \models K_B\,(C\clubsuit \;\wedge K_D\neg K_B C\clubsuit)$

Read K_D as D believes, rather than D knows.

What mathematical modeling is all about



We'll call them Amina (A) (female) and Bao (B) (male).

A and *B* enter a large room containing a remote-control mechanical coin flipper.
One presses a button, and the coin spins through the air, landing in a small box on a table.

The box closes.

The two people are much too far to see the coin.

In reality, the coin shows heads.

Two people enter a room: intuitions

The coin shows heads.

The coin doesn't show tails.

A doesn't know that the coin shows heads.

B doesn't know that the coin shows heads.

A doesn't know that the coin shows tails.

B doesn't know that the coin shows tails.

A knows that she doesn't know that the coins shows heads.

B knows that he doesn't know that the coins shows heads.

A knows that *B* doesn't know that the coins shows heads.

B knows that *A* doesn't know that the coins shows heads.

The coin shows heads.

The coin doesn't show tails.

A doesn't know that the coin shows heads.

B doesn't know that the coin shows heads.

A doesn't know that the coin shows tails.

B doesn't know that the coin shows tails.

A knows that she doesn't know that the coins shows heads.

B knows that he doesn't know that the coins shows heads.

A knows that *B* doesn't know that the coins shows heads.

B knows that *A* doesn't know that the coins shows heads.

ASSUMPTION

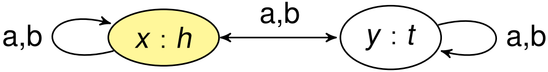
We all agree that these are reasonable intuitions.

GOING ON

We'll be interested in formalizing these, and then in working with the formalizations.

The coin shows heads.
The coin doesn't show tails.
A doesn't know that the coin shows heads.
B doesn't know that the coin shows heads.
A doesn't know that the coin shows tails.
B doesn't know that the coin shows tails.
A knows that she doesn't know that the coins shows heads.
B knows that he doesn't know that the coins shows heads.
A knows that *B* doesn't know that the coins shows heads.
B knows that *A* doesn't know that the coins shows heads.

Let's use atomic sentences h and t.
The coin shows heads. h
The coin doesn't show tails. $\neg t$
A doesn't know that the coin shows heads. $\neg K_a h$
B doesn't know that the coin shows heads. $\neg K_b h$
A doesn't know that the coin shows tails. $\neg K_a t$
B doesn't know that the coin shows tails. $\neg K_b h$
A knows that she doesn't know that the coins shows heads.
$K_a \neg K_a h$
B knows that he doesn't know that the coins shows heads.
$K_b \neg K_b h$
A knows that B doesn't know that the coins shows heads.
$K_a \neg K_b h$
B knows that A doesn't know that the coins shows heads.
$K_b \neg K_a h$



The real world x is shaded yellow.

What the picture is supposed to represent

The picture is a proposal for the intersubjective reality shared between A and B, the reality that is the basis of their claims of knowledge and ignorance.

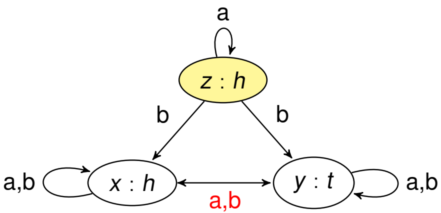
THE DEFINITIONS

$x \models p$	iff	p is written on x (for p atomic)
$x \models \varphi \wedge \psi$	iff	$x \models \varphi$ and $x \models \psi$
$x \models \varphi \vee \psi$	iff	$x \models \varphi$ or $x \models \psi$
$x \models \neg\varphi$	iff	if it is not the case that $x \models \varphi$
$x \models \varphi \to \psi$	iff	if $x \models \varphi$, then $x \models \psi$
$x \models \varphi \leftrightarrow \psi$	iff	$x \models \varphi$ if and only if $x \models \psi$
$x \models K_b\, \varphi$	iff	$y \models \varphi$ for all y such that $x \to y$ for b

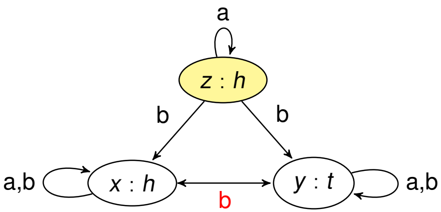
AS A RESULT

$x \not\models p$	iff	p is not written on x (for p atomic)
$x \not\models \varphi \wedge \psi$	iff	$x \not\models \varphi$ or $x \not\models \psi$
$x \not\models \varphi \vee \psi$	iff	$x \not\models \varphi$ and $x \not\models \psi$
$x \not\models \neg\varphi$	iff	$x \models \varphi$
$x \not\models \varphi \to \psi$	iff	$x \models \varphi$, but $x \not\models \psi$
$x \not\models \varphi \leftrightarrow \psi$	iff	($x \models \varphi$, but $x \not\models \psi$) or ($x \models \psi$, but $x \not\models \varphi$)
$x \not\models K_b\, \varphi$	iff	for some y such that $x \to y$ for b, $y \not\models \varphi$

Why



and not



Comment on the model for "cheating"

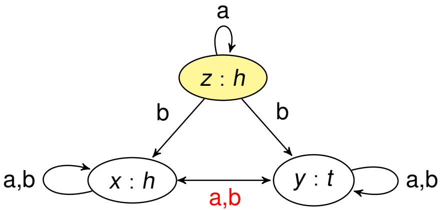
Consider the sentence
B knows/believes that *A* doesn't know that the coin lies heads up

This sentence is intutively true about the cheating scenario.

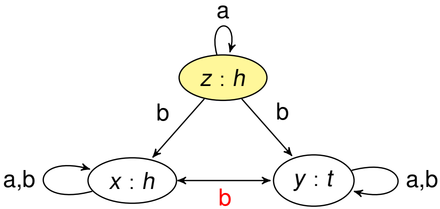
The formal version is

$$K_b \neg K_a \ h$$

Model 1



Model 2



In Model 1, $z \models K_b \neg K_a \; h$

In Model 2, $z \not\models K_b \neg K_a \; h$

After Amina sneaks a peak, Bao does the same thing.
Neither even think it is possible that the other one looked.

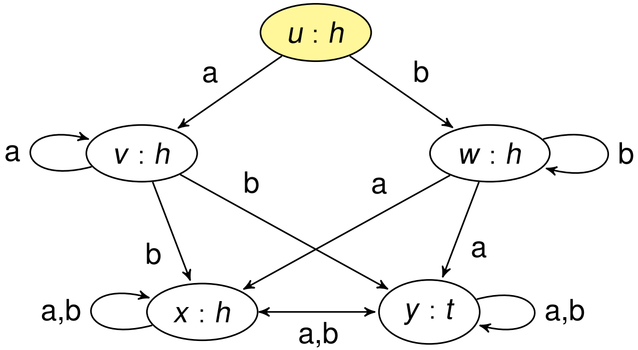
Try to draw the model, indicating the real world.

How to evaluate models

The real test of a proposed model is
to answer the following question:

Are all of our intuitions truth or falsity of English sentences
about the scenario
matched by
truth or falsity of logic sentences
in the real world of the model?

After Amina sneaks a peak, Bao does the same thing.
Neither even think it is possible that the other one looked.



This can be checked just as we checked the previous models: by formalizing intuitions and checking them at the "real world" *u*.

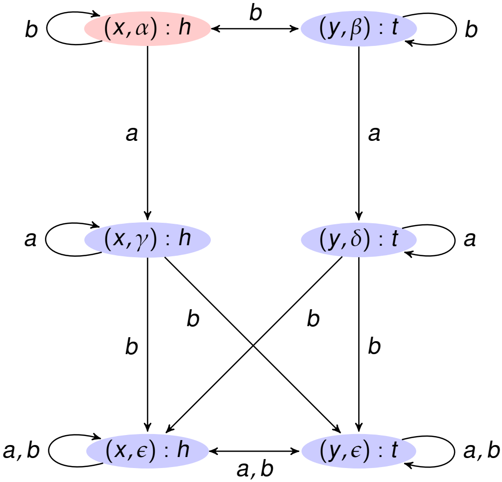
After the first coin scenario, *A* secretly opens the box herself.

B does not observe *A* open the box,
A is certain that *B* did not suspect that anything happened at all.

But in reality, *B* did see *A* open the box (but not the coin itself).

Try to draw the model for this.

The product is then



The point is that one can now ask if this model matches our intuitions.

Let S be a real-world scenario involving people and some set of well-understood sentences about S.
Then there is a finite set T of natural language sentences which explains S in a precise enough manner so that all possible statements about knowledge, higher-order knowledge, and common knowledge have determined truth values.

Furthermore, corresponding to S is a mathematical object, a finite multi-agent Kripke model $\hat{S}$, a world $s_0 \in \hat{S}$, and a set $\hat{T}$ of sentences in the formal language of multi-agent epistemic logic $\mathcal{L}$ such that for each natural language sentence A about the same atomic facts and the same participants in S, the following are equivalent:

1. A is intuitively true about S.
2. $s_0 \models \hat{A}$ in $\hat{S}$.

PART III: Meet Ronnie, the Logical Rat

Ronnie is a very smart rat. He can do logic!

He is the mascot of a company that manufactures two-chambered rat houses.

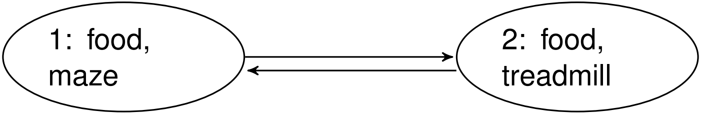
Each day, they put him in a different house.
He runs back and forth all day long.

PART III: Meet Ronnie, the Logical Rat

Ronnie is a very smart rat. He can do logic!

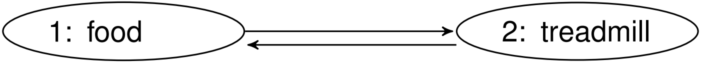
At any given time, Ronnie is very intested in what he finds in the room he currently is in,
and also what he sees in the room he is *not* in.

Although the rooms are numbered from the outside,
he doesn't "know" about numbers.



So in room 1, there is food and a maze (and no treadmill)
In room 2, there is food and a treadmill (and no maze).

Here is a different model:



So in room 1, there is food (and no treadmill).
In room 2, there is a treadmill (and no food).

He's a good reasoner, and so he needs a logical language in which to make assertions.
The atomic sentences of this language should correspond to the presence of objects familiar to Ronnie.
We'll write these atomic sentences as

$$p_1, p_2, p_3, \ldots. p_n, \ldots$$

Each of these corresponds to some item.
The set AtSen is the set of all atomic sentences.
For example, p_1 might correspond to a water dish,
p_2 to a treadmilll,
p_3 to a maze, etc.

Note that in a given house,
p_n might be true in one chamber and not the other,
or in both, or in neither.

So far in this tutorial we have seen
epistemic logic (for dealing with knowledge)
and the bare-bones propositional (or sentential) logic
which we'll write as $\mathcal{L}$.

This language is $\mathcal{L}$ not expressive enough for Ronnie:
p_1 means that there is a water dish *here*,
and he might want to say that there is a water dish *there*.

We add a new operator to $\mathcal{L}$ called $*$, which we read as “switch”.

Now we get a new language which we call $\mathcal{L}(*)$.

So Ronnie would use $*p_1$ to say that there is a water dish in the chamber that he is *not* currently occupying.

This language $\mathcal{L}(*)$ also has complex sentences in it like

$$*(p_2 \to *\neg p_3).$$

This would say that in the other room from where he currently is,

from the point of view of the other room,
if there is a treadmill [in that room]
then if I would switch rooms [getting back to *here*],
there would be no maze

We are given a model *M*.
This is just a pair of worlds, called 1 and 2.

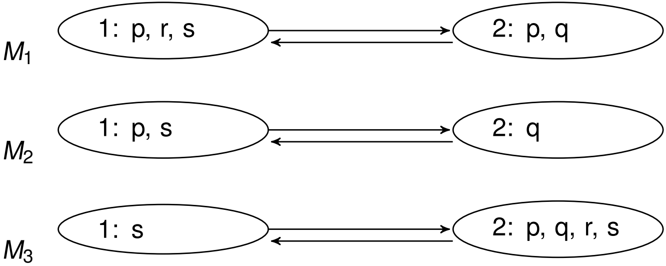
The model also comes with
a statement of which atomic sentences are
are true in world 1, and which are true in world 2.

Mathematically, a model for $\mathcal{L}(*)$ is a pair of sets of atomic sentences.
We prefer to draw them as graphs, but technically, they're pairs of sets.

We are given a model *M*.
This is just a pair of worlds, called 1 and 2.

The model also comes with
a statement of which atomic sentences are
are true in world 1, and which are true in world 2.

Mathematically, a model for $\mathcal{L}(*)$ is a pair of sets of atomic sentences.
We prefer to draw them as graphs, but technically, they're pairs of sets.



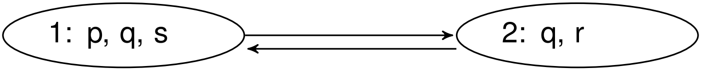
We are given a model *M*.
This is just a pair of worlds, called 1 and 2.

The model also comes with
a statement of which atomic sentences are
are true in world 1, and which are true in world 2.

Mathematically, a model for $\mathcal{L}(*)$ is a pair of sets of atomic sentences.
We prefer to draw them as graphs, but technically, they're pairs of sets.

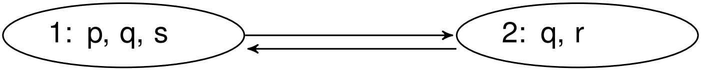
We then define:

$1 \Vdash p$	iff	if the model says so	$2 \Vdash p$	iff	if the model says so
$1 \Vdash \neg\varphi$	iff	not $(1 \Vdash \varphi)$	$2 \Vdash \neg\varphi$	iff	not $(2 \Vdash \varphi)$
$1 \Vdash \varphi \wedge \psi$	iff	$1 \Vdash \varphi$ and $1 \Vdash \psi$	$2 \Vdash \varphi \wedge \psi$	iff	$2 \Vdash \varphi$ and $1 \Vdash \psi$
$1 \Vdash \varphi \vee \psi$	iff	$1 \Vdash \varphi$ or $1 \Vdash \psi$	$2 \Vdash \varphi \vee \psi$	iff	$2 \Vdash \varphi$ or $1 \Vdash \psi$
$1 \Vdash *\varphi$	iff	$2 \Vdash \varphi$	$2 \Vdash *\varphi$	iff	$1 \Vdash \varphi$



φ	$1 \Vdash \varphi$	$2 \Vdash \varphi$
p		
q		
$*r$		
$p \wedge *q$		
$*(p \wedge *q)$		
$*(p \wedge *q) \rightarrow *r$		

Fill in the table with √ for Yes and × for No.



φ	$1 \Vdash \varphi$	$2 \Vdash \varphi$
p	$\surd$	$\times$
q	$\surd$	$\surd$
$*r$	$\surd$	$\times$
$p \wedge *q$	$\surd$	$\times$
$*(p \wedge *q)$	$\times$	$\surd$
$*(p \wedge *q) \rightarrow *r$	$\surd$	$\times$

$\surd$ means Yes and $\times$ means No.

Ronnie wants a logical system which allows him to do general logical reasoning.
Specifically, he wants a system which should correspond to a natural semantic notion of validity.
The trouble is that he has come up with not one, not two, but three notions of validity:

$\models_1 \varphi$	iff	for all models M, $1 \Vdash \varphi$ in M
$\models_2 \varphi$	iff	for all models M, $2 \Vdash \varphi$ in M
$\models_{1,2} \varphi$	iff	for all models M, $1 \Vdash \varphi$ in M, and also $2 \Vdash \varphi$ in M

Intuitively,

$\models_1 \varphi$	means	no matter which rat house Ronnie is in he can be sure that φ is true in room 1
$\models_2 \varphi$	means	no matter which rat house Ronnie is in he can be sure that φ is true in room 2
$\models_{1,2} \varphi$	means	no matter which rat house Ronnie is in he can be sure that φ is true in both rooms

$\models_1 \varphi$	iff	for all models M, $1 \Vdash \varphi$ in M
$\models_2 \varphi$	iff	for all models M, $2 \Vdash \varphi$ in M
$\models_{1,2} \varphi$	iff	for all models M, $1 \Vdash \varphi$ in M, and also $2 \Vdash \varphi$ in M

Fortunately, all three notions shown above turn out to be the same.

THEOREM: FOR ALL φ, $\models_1 \varphi$ IFF $\models_2 \varphi$ IFF $\models_{1,2} \varphi$.

But this is a special feature of this setting!

$\models_1 \varphi$	iff	for all models M, $1 \Vdash \varphi$ in M
$\models_2 \varphi$	iff	for all models M, $2 \Vdash \varphi$ in M
$\models_{1,2} \varphi$	iff	for all models M, $1 \Vdash \varphi$ in M, and also $2 \Vdash \varphi$ in M

However, the equivalence no longer holds when we allow assumptions on the left.

$\varphi_1, \ldots, \varphi_n \models_1 \psi$	iff	for all models M, such that in M, $1 \Vdash \varphi_1$, …, $1 \Vdash \varphi_n$, we have $1 \Vdash \psi$ in M as well
$\varphi_1, \ldots, \varphi_n \models_2 \psi$	iff	for all models M, such that in M, $2 \Vdash \varphi_1$, …, $2 \Vdash \varphi_n$, we have $2 \Vdash \psi$ in M as well
$\varphi_1, \ldots, \varphi_n \models_{1,2} \psi$	iff	for all models M, such that in M, $1 \Vdash \varphi_1$, …, $1 \Vdash \varphi_n$, and also $2 \Vdash \varphi_1$, …, $2 \Vdash \varphi_n$, we have $1 \Vdash \psi$ and also $2 \Vdash \psi$ as well

If we take $n = 0$, then the φ's go away,
the "such that" phrases disappear,
and we get our old definitions back of $\models_1 \varphi$, $\models_2 \varphi$, and $\models_{1,2} \varphi$.

Lemma: The following are equivalent:

1. $\varphi_1, \ldots, \varphi_n \models_1 \psi$
2. $\varphi_1, \ldots, \varphi_n \models_2 \psi$

This is a good exercise.

We call $\models_1$ the local consequence relation,
and $\models_{1,2}$ the global consequence relation.

The two agree when $n = 0$, as we know:

$$\models_1 \psi \quad \text{iff} \quad \models \psi.$$

One of these is true, one is false. Which is which?

$$\varphi_1, \ldots, \varphi_n \models_1 \psi \quad \text{iff} \quad \models (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \psi.$$

$$\varphi_1, \ldots, \varphi_n \models_{1,2} \psi \quad \text{iff} \quad \models (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \psi.$$

We are interested in derivations with no “premise” lines.
First, we take two axiom schemes

Involution axioms $**\varphi \to \varphi$.

Determinacy axioms $\neg * \varphi \to *\neg\varphi$.

Then we adopt a new type of subproof, called a $*$-subproof.
At any point, one may start a $*$-subproof.

Sentences which begin with a $*$ or $\neg*$ may be copied into a $*$-subproof.
And then when one is copied in, the outermost $*$ is dropped ($*$e).
A $*$-subproof may be closed at any time, and then its last line may be copied out, but with a $*$ added ($*$i).

We’ll use $\perp$ as a symbol for contradiction reached.

Example: $\vdash (**\mathit{food} \wedge ***\mathit{water}) \rightarrow *(\mathit{water} \wedge *\mathit{food})$.

We use the following derivation:

1		$**\mathit{food} \wedge ***\mathit{water}$	assumption
2	*	$*\mathit{food}$	$\wedge$ e, $*$ e, 1
3		$**\mathit{water}$	$\wedge$e, $*$ e, 1
4		$**\mathit{water} \rightarrow \mathit{water}$	involution axiom
5		water	$\rightarrow$e, 4, 3
6		$\mathit{water} \wedge *\mathit{food}$	$\wedge$i, 5, 2
7		$*(\mathit{water} \wedge *\mathit{food})$	$*$i, 2–6
8		$(**\mathit{food} \wedge ***\mathit{water}) \rightarrow *(\mathit{water} \wedge *\mathit{food})$	$\rightarrow$i, 1–7

The proof begins with an assertion that we assume about one of the two rooms *i* in some arbitrary rat house *M*.
Line 2 begins a $*$-subproof, indicating that we shift focus to the other room $3 - i$.
In that room, the conjuncts in line 1 are true, provided that we drop the outermost $*$.
One of these statements is $**$ *water*.
Using an Involution Axiom, we see that *water* is indeed true in our current room.
We continue to reason about room $3 - i$ until line 6, and then we pop back to room *i*.
This shift in view is the reason we add a $*$ to line 6 when we wrote it in line 7.

- The logic is dynamic in that the point of semantic evaluation may change due to operators in the syntax.
- We could also add operators for ontic effects: *drink the water!*.
- The notion of validity has a subtlety: local vs. global.
- Proof systems are possible, with varying properties.

One of the most important group-level knowledge phenomena is common knowledge.

A sentence S is common knowledge for a group $\mathcal{G}$ of people if

- ⋆ Everyone in $\mathcal{G}$ knows S.
- ⋆ Everyone in $\mathcal{G}$ knows that everyone in $\mathcal{G}$ knows S.
- ⋆ Everyone in $\mathcal{G}$ knows that everyone in $\mathcal{G}$ knows that everyone in $\mathcal{G}$ knows S.
- ⋆ etc., forever.

We hasten to add that this is an informal definition;
very soon we'll put forward a proposal on how this might be modeled with modal logic.

We have already seen the most basic "coin scenario."

> A and B enter a large room containing a remote-control mechanical coin flipper.
> One presses a button, and the coin spins through the air, landing in a small box on a table.
> The box closes.
> A and B are much too far to see the coin.
> In reality, the coin shows heads.

The group $\mathcal{G}$ here is the set $\{A, B\}$.

It should be common knowledge for this group that neither knows the state of the coin.

From our introduction:

We have a deck with three cards ♥, ♣, and ♦. We also have three players: B, C, and D. We deal the cards out, one to each player. The deal was $(B : ♦, C : ♣, D : ♥)$. Then, C and D show each other their cards while B looks on.

Let us consider the group $\mathcal{G} = \{C, D\}$.
It should be common knowledge for this group that C has ♣, D has ♥, and that B doesn't know any of this.

But if we considered the larger group $\{B, C, D\}$, then it is not common knowledge that C has ♣.

Countries differ as to which side of the road one drives a car;
the matter is one of social and legal convention.
In Kenya, they follow British custom and drive on the left.

Suppose that in Kenya, the government decides to change the driving side.

But suppose that the change is made in a quiet way,
so that only one person in the country, say Silvanos,
finds out about it.
After this, what should Silvanos do?

Countries differ as to which side of the road one drives a car;
the matter is one of social and legal convention.
In Kenya, they follow British custom and drive on the left.

Suppose that in Kenya, the government decides to change the driving side.

But suppose that the change is made in a quiet way,
so that only one person in the country, say Silvanos,
finds out about it.
After this, what should Silvanos do?

From the point of view of safety, it is clear that
he should not obey the law:
since others will be disobeying it, he puts his life at risk.

Digression: common knowledge and social conventions

Suppose further that the next day the government decides to make an announcement to the press that the law was changed. What should happen now?

Suppose further that the next day the government decides to make an announcement to the press that the law was changed. What should happen now?

The streets are more dangerous and more unsure this day, because many people will still not know about the change. Even the ones that have heard about it will be hesitant to change, since they do not know whether the other drivers know or not.

Digression: common knowledge and social conventions

Eventually, after further announcements, we reach a state where:

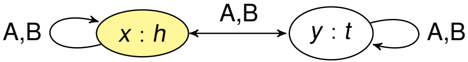
$$\text{The law says drive on the right and everyone knows (1).} \quad (1)$$

Note that (1) is a circular statement. The key point is not that everyone know what the law says, but that they in addition know this very fact, the content of the sentence you are reading.

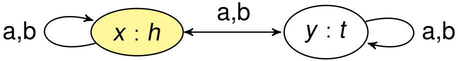
A group $\mathcal{G}$ of agents is a non-empty subset of $\mathcal{A}$.
As with all set notation, the order in which elements of a set are listed doesn't count, and neither do repeats.

In this tutorial, I use both
upper-case and lower case letters for the agents.

I hope you don't mind this inconsistence.
For example, the coin scenario could be rendered as either



or equally well as



We are building on epistemic logic as we studied it earlier this semester.
So we have a set $\mathcal{A}$ of agents, and a set AtSen of atomic sentences,
and we make new sentences using

- ⋆ the connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$
- ⋆ the modal operators K_A for each agent A.

For each group $\mathcal{G}$ of agents, we add two new operators.

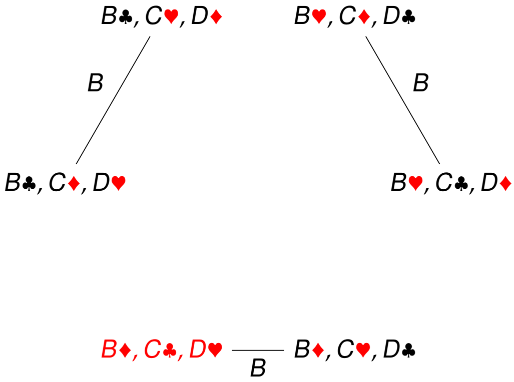
- ⋆ $E_{\mathcal{G}}$, so that $E_{\mathcal{G}}\,\varphi$ represents "everyone in $\mathcal{G}$ knows that φ."
- ⋆ $CK_{\mathcal{G}}$, so that $CK_{\mathcal{G}}\,\varphi$ represents "it is common knowledge for the group $\mathcal{G}$ that φ."

Here is our the formal definition of the semantics of $E_{\mathcal{G}}\,\varphi$.
We start with a multi-agent model $\mathcal{M} = (W, (R_a)_{A\in\mathcal{A}}, val)$.
For a world $x \in W$,

$$x \Vdash E_{\mathcal{G}}\,\varphi \quad \text{iff} \quad \begin{array}{l}\text{for all } A \in \mathcal{G} \\ \text{and all worlds } y \text{ such that } x\ R_A\ y, \\ y \Vdash \varphi\end{array}$$

Semantics of $E_{\mathcal{G}}\,\varphi$: an example

We return to the three-card scenario earlier in this tutorial.



(The self-loops on all the nodes are omitted.)

B♦, C♣, D♥ $\Vdash E_{\{C,D\}} C$♣.
B♦, C♣, D♥ $\Vdash E_{\{C,D\}} \neg K_B\ C$♣.
B♦, C♣, D♥ $\Vdash \neg E_{\{B,C,D\}} C$♣.

The main sound logical principle for the $E_{\mathcal{G}}$ operators is:

$$\models_{\text{all}} E_{\mathcal{G}}\,\varphi \leftrightarrow \bigwedge_{A \in \mathcal{G}} K_A\ \varphi.$$

For example

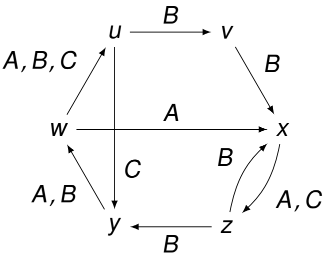
$$\models_{\text{all}} E_{\{B,C\}}\varphi \leftrightarrow (K_B\ \varphi\ \wedge\ K_C\ \varphi)$$

$$\models_{\text{all}} E_{\{A,B,C\}}\varphi \leftrightarrow (K_A\ \varphi\ \wedge\ K_B\ \varphi\ \wedge\ K_C\ \varphi)$$

For a one-element set such as $\{B\}$,
$E_{\{B\}}\varphi$ would be literally the same as $K_B\varphi$.

Preliminary definition: Given a multi-agent model $\mathcal{M} = (W, (R_a)_{A\in\mathcal{A}}, val)$ and a group $\mathcal{G} \subseteq \mathcal{A}$:

$x \xrightarrow{\mathcal{G}^+} y$ iff there is a sequence of length ≥ 1 from x to y using arrows labeled by agents in the set $\mathcal{G}$



Yes:	$u \xrightarrow{\{B,C\}^+} u$	$w \xrightarrow{\{A,C\}^+} z$	$y \xrightarrow{\{B\}^+} x$
No:	$u \xrightarrow{\{B\}^+} u$	$w \xrightarrow{\{A\}^+} w$	$z \xrightarrow{\{A,C\}^+} y$

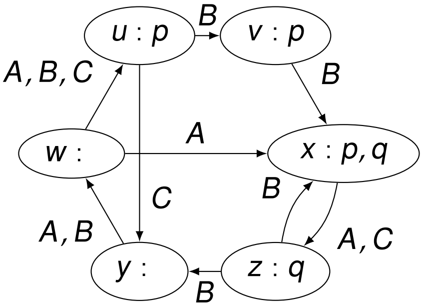
Preliminary definition: Given a multi-agent model $\mathcal{M} = (W, (R_a)_{A \in \mathcal{A}}, val)$ and a group $\mathcal{G} \subseteq \mathcal{A}$:

$x \xrightarrow{\mathcal{G}^+} y$ IFF THERE IS A SEQUENCE OF LENGTH ≥ 1 FROM x TO y USING ARROWS LABELED BY AGENTS IN THE SET $\mathcal{G}$

Here is our the formal definition of the semantics of $CK_{\mathcal{G}}\,\varphi$.
We start with a multi-agent model $\mathcal{M} = (W, (R_a)_{A \in \mathcal{A}}, val)$.
For a world $x \in W$,

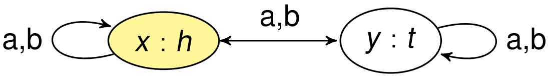
$x \Vdash CK_{\mathcal{G}}{}^{+}\varphi$ iff for all y such that $x \xrightarrow{\mathcal{G}^+} y$, $y \Vdash \varphi$

$x \Vdash CK_{\mathcal{G}}\,\varphi$ iff for all y such that $x = y$ or $x \xrightarrow{\mathcal{G}^+} y$, $y \Vdash \varphi$



Yes: $u \Vdash CK_{\{B\}}\, p$ $\quad x \Vdash CK_{\{A,C\}}\, q$ $\quad z \Vdash CK_{\{B\}}(q \rightarrow p)$

No: $u \Vdash CK_{\{B\}}\, p$ $\quad w \Vdash CK_{\{A\}}\, p$ $\quad z \Vdash CK_{\{A,B,C\}}(q \rightarrow p)$



$x \Vdash CK_{A,B}(h \oplus t)$

that is, $x \Vdash CK_{A,B}((h \wedge \neg t) \vee (\neg h \wedge t))$.

and also $x \Vdash CK_{A,B} \neg K_a h$.

The statements which are common knowledge typically formalize “hard facts” which we bring to various social situations.

If we have a "small" group of agents, such as $\{a, b\}$, then often we drop the set braces from the notation.

We might write $CK_{a,b}\ \varphi$ instead of $CK_{\{a,b\}}\ \varphi$.

In fact, the set braces on this look a little pedantic.

The main sound logical principles for the $CK_{\mathcal{G}}$ operators are: the Mix Axiom

$$CK_{\mathcal{G}}\,\varphi \to \varphi \wedge E_{\mathcal{G}}\,CK_{\mathcal{G}}\,\varphi$$

(so-called because it deals with the interactions of the two operators $CK_{\mathcal{G}}$ and $E_{\mathcal{G}}$)
Also, we have the Induction Rule:

from $\chi \to \psi \wedge K_a\chi$ for all $a \in \mathcal{G}$, infer $\chi \to CK_{\mathcal{G}}\,\psi$.

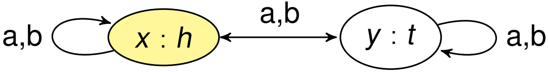
Axioms	All tautologies	
	(Normality)	$K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$
	(Veracity)	$K_a\,\varphi \to \varphi$
	(Positive Introspection)	$K_a\,\varphi \to K_a\,K_a\,\varphi$
	(Negative Introspection)	$\neg K_a\,\varphi \to K_a\,\neg K_a\,\varphi$
	(Everyone)	$E_{\mathcal{G}}\,\varphi \leftrightarrow \bigwedge_{a\in\mathcal{A}} K_a\varphi$
	(Mix)	$CK_{\mathcal{G}}\,\varphi \to \varphi \wedge E_{\mathcal{G}}\,(\varphi \wedge CK_{\mathcal{G}}\,\varphi)$
Rules	(Modus Ponens)	From φ and $\varphi \to \psi$, infer ψ
	(Necessitation)	From φ, infer $K_a\varphi$
	(Induction)	From $\chi \to K_a\chi$ for all $a \in \mathcal{G}$, infer $\chi \to CK_{\mathcal{G}}\,\psi$

Using this logic, one can prove the important properties of common knowledge. For example, the transitivity property:

$$CK_{\mathcal{G}}\,\varphi \to CK_{\mathcal{G}}\,CK_{\mathcal{G}}\,\varphi.$$

If φ is common knowledge in a group, then the fact of its being common knowledge is itself common knowledge in the group.

Formalize three intuitive truths about the first coin scenario



Formalize these as φ_1, φ_2, and φ_3.
Then show that the sentences φ_i are all provable consequences of

$$h \;\wedge\; CK_{\{a,b\}}((h \oplus t) \;\wedge\; \neg K_a h \;\wedge \neg K_a t \;\wedge \neg K_b h \;\wedge \neg K_b t)$$

That is, call the sentence above χ.
Show about the formalized sentences φ_1, φ_2, and φ_3 that

$$\vdash \chi \rightarrow \varphi_1$$

and similarly for the other two.

In New Brunswick, let's say that restaurant meals are either

Good or Bad

and that we want to rate a restaurant as

$\star\star\star\star\star$ or $\star$

Please note that there is a big difference between
our private experience (what we eat)
and our public rating.

In what follows, our rating is allowed to
depend on the ratings of other people,
but not their experience.

This discussion has several goals and meta-goals:

What do you think they are?

Let's make some assumptions.

- 50% of the restaurants in town are ★★★★★
  50% of the restaurants in town are ★.
- At a ★★★★★ bistro,
  66% of the meals are good, and 34% are bad.
- At a ★ joint,
  34% of the meals are good, and 66% are bad.

It's very hard to give a model of how actual people will fare in this kind of thing.

So we are going to step back and do something which is in some ways similar,
and in some ways different.

We are going to assume that
everyone involved is attending NASSLLI, and thus (?!) know and uses probability.

So given some decision, everyone will use probability (and thus logic)
to "weigh all the evidence."

Moreover, everyone involved knows that everyone else knows and uses probability and is behaving the same way,
etc.

$\mathbb{R}$ is the set of real numbers, and $\mathbb{R}^{\geq 0}$ is the set of non-negative reals.

A *probability space* is a set S and a function $\Pr : S \to \mathbb{R}^{\geq 0}$

$$\sum_{s \in S} \Pr(s) \quad = \quad 1.$$

Events

An event is a subset $A \subseteq S$.

For example, $\emptyset$ and S are events.

We define

$$\Pr(A) \quad = \quad \sum_{s \in A} \Pr(s).$$

So $\Pr(\emptyset) = 0$ and $\Pr(S) = 1$.

We have a probability space $(S, \Pr)$.
Also functions $X : S \to V$ for some V. (Usually $V = \mathbb{R}$.)
Let the image of X be $x_1, \ldots, x_n$.
X is called a random variable.
Usually S is implicit.

Note on the terminology

Calling X a “random” variable
might be mistleading.

There doesn’t have to be anything “random” about it.

A random variable X induces a partition of S: for x_i we have

$$X = x_i \quad = \quad \{s \in S : X(s) = x_i\}$$

and we also have $Pr[X = x_i]$.
(These are standard kinds of notation.
The important point is that $X = x_i$ is an event.)

Let A and B be events in the same probability space.

The conditional probability of A given B is defined by:

$$\Pr(A \mid B) \quad = \quad \frac{\Pr(A \cap B)}{\Pr(B)}$$

We only use this notation when $\Pr(B) \neq 0$.

The first sample space:

We have a set with two outcomes

$$\{\star\star\star\star\star, \star\}$$

and $Pr(\star^5) = .5$, $Pr(\star) = .5$.

This is our probability estimates of randomly-chosen restaurants
with no experience of our own and no ratings.

These calculations are with Pr(five star) = .5, Pr(one star) = .5, Pr(good | five star) = .66, Pr(bad | five star) = .34, Pr(good | one star) = .34, Pr(bad | one star) = .66.

number	state	probability
0	[five_star, good, good, good, good, good]	0.062617
1	[five_star, good, good, good, good, bad]	0.032257
2	[five_star, good, good, good, bad, good]	0.032257
3	[five_star, good, good, good, bad, bad]	0.016617
4	[five_star, good, good, bad, good, good]	0.032257
5	[five_star, good, good, bad, good, bad]	0.016617
6	[five_star, good, good, bad, bad, good]	0.016617
7	[five_star, good, good, bad, bad, bad]	0.0085604
8	[five_star, good, bad, good, good, good]	0.032257
9	[five_star, good, bad, good, good, bad]	0.016617
10	[five_star, good, bad, good, bad, good]	0.016617
11	[five_star, good, bad, good, bad, bad]	0.0085604
12	[five_star, good, bad, bad, good, good]	0.016617
13	[five_star, good, bad, bad, good, bad]	0.0085604
14	[five_star, good, bad, bad, bad, good]	0.0085604
15	[five_star, good, bad, bad, bad, bad]	0.0044099
16	[five_star, bad, good, good, good, good]	0.032257
17	[five_star, bad, good, good, good, bad]	0.016617
18	[five_star, bad, good, good, bad, good]	0.016617
19	[five_star, bad, good, good, bad, bad]	0.0085604
20	[five_star, bad, good, bad, good, good]	0.016617
21	[five_star, bad, good, bad, good, bad]	0.0085604
22	[five_star, bad, good, bad, bad, good]	0.0085604
23	[five_star, bad, good, bad, bad, bad]	0.0044099
24	[five_star, bad, bad, good, good, good]	0.016617
25	[five_star, bad, bad, good, good, bad]	0.0085604
26	[five_star, bad, bad, good, bad, good]	0.0085604
27	[five_star, bad, bad, good, bad, bad]	0.0044099
28	[five_star, bad, bad, bad, good, good]	0.0085604
29	[five_star, bad, bad, bad, good, bad]	0.0044099
30	[five_star, bad, bad, bad, bad, good]	0.0044099
31	[five_star, bad, bad, bad, bad, bad]	0.0022718

number	state	algorithm
32	[one_star, good, good, good, good, good]	0.0022718
33	[one_star, good, good, good, good, bad]	0.0044099
34	[one_star, good, good, good, bad, good]	0.0044099
35	[one_star, good, good, good, bad, bad]	0.0085604
36	[one_star, good, good, bad, good, good]	0.0044099
37	[one_star, good, good, bad, good, bad]	0.0085604
38	[one_star, good, good, bad, bad, good]	0.0085604
39	[one_star, good, good, bad, bad, bad]	0.016617
40	[one_star, good, bad, good, good, good]	0.0044099
41	[one_star, good, bad, good, good, bad]	0.0085604
42	[one_star, good, bad, good, bad, good]	0.0085604
43	[one_star, good, bad, good, bad, bad]	0.016617
44	[one_star, good, bad, bad, good, good]	0.0085604
45	[one_star, good, bad, bad, good, bad]	0.016617
46	[one_star, good, bad, bad, bad, good]	0.016617
47	[one_star, good, bad, bad, bad, bad]	0.032257
48	[one_star, bad, good, good, good, good]	0.0044099
49	[one_star, bad, good, good, good, bad]	0.0085604
50	[one_star, bad, good, good, bad, good]	0.0085604
51	[one_star, bad, good, good, bad, bad]	0.016617
52	[one_star, bad, good, bad, good, good]	0.0085604
53	[one_star, bad, good, bad, good, bad]	0.016617
54	[one_star, bad, good, bad, bad, good]	0.016617
55	[one_star, bad, good, bad, bad, bad]	0.032257
56	[one_star, bad, bad, good, good, good]	0.0085604
57	[one_star, bad, bad, good, good, bad]	0.016617
58	[one_star, bad, bad, good, bad, good]	0.016617
59	[one_star, bad, bad, good, bad, bad]	0.032257
60	[one_star, bad, bad, bad, good, good]	0.016617
61	[one_star, bad, bad, bad, good, bad]	0.032257
62	[one_star, bad, bad, bad, bad, good]	0.032257
63	[one_star, bad, bad, bad, bad, bad]	0.062617

The upshot, assuming as we do that the first person follows the social procedure

The first person believes that a good meal indicates $\star^5$,
and that a bad meal indicates $\star$.

These calculations are with $Pr(\texttt{five star}) = .5$, $Pr(\texttt{one star}) = .5$, $Pr(\texttt{good} \mid \texttt{five star}) = .66$, $Pr(\texttt{bad} \mid \texttt{five star}) = .34$, $Pr(\texttt{good} \mid \texttt{one star}) = .34$, $Pr(\texttt{bad} \mid \texttt{one star}) = .66$.

number	state	algorithm
0	[good, good, good, good, good]	[five_star, five_star, five_star, five_star, five_star]
1	[good, good, good, good, bad]	[five_star, five_star, five_star, five_star, five_star]
2	[good, good, good, bad, good]	[five_star, five_star, five_star, five_star, five_star]
3	[good, good, good, bad, bad]	[five_star, five_star, five_star, five_star, five_star]
4	[good, good, bad, good, good]	[five_star, five_star, five_star, five_star, five_star]
5	[good, good, bad, good, bad]	[five_star, five_star, five_star, five_star, five_star]
6	[good, good, bad, bad, good]	[five_star, five_star, five_star, five_star, five_star]
7	[good, good, bad, bad, bad]	[five_star, five_star, five_star, five_star, five_star]
8	[good, bad, good, good, good]	[five_star, one_star, five_star, five_star, five_star]
9	[good, bad, good, good, bad]	[five_star, one_star, five_star, five_star, five_star]
10	[good, bad, good, bad, good]	[five_star, one_star, five_star, one_star, five_star]
11	[good, bad, good, bad, bad]	[five_star, one_star, five_star, one_star, one_star]
12	[good, bad, bad, good, good]	[five_star, one_star, one_star, five_star, five_star]
13	[good, bad, bad, good, bad]	[five_star, one_star, one_star, five_star, one_star]
14	[good, bad, bad, bad, good]	[five_star, one_star, one_star, one_star, one_star]
15	[good, bad, bad, bad, bad]	[five_star, one_star, one_star, one_star, one_star]
16	[bad, good, good, good, good]	[one_star, five_star, five_star, five_star, five_star]
17	[bad, good, good, good, bad]	[one_star, five_star, five_star, five_star, five_star]
18	[bad, good, good, bad, good]	[one_star, five_star, five_star, one_star, five_star]
19	[bad, good, good, bad, bad]	[one_star, five_star, five_star, one_star, one_star]
20	[bad, good, bad, good, good]	[one_star, five_star, one_star, five_star, five_star]
21	[bad, good, bad, good, bad]	[one_star, five_star, one_star, five_star, one_star]
22	[bad, good, bad, bad, good]	[one_star, five_star, one_star, one_star, one_star]
23	[bad, good, bad, bad, bad]	[one_star, five_star, one_star, one_star, one_star]
24	[bad, bad, good, good, good]	[one_star, one_star, one_star, one_star, one_star]
25	[bad, bad, good, good, bad]	[one_star, one_star, one_star, one_star, one_star]
26	[bad, bad, good, bad, good]	[one_star, one_star, one_star, one_star, one_star]
27	[bad, bad, good, bad, bad]	[one_star, one_star, one_star, one_star, one_star]
28	[bad, bad, bad, good, good]	[one_star, one_star, one_star, one_star, one_star]
29	[bad, bad, bad, good, bad]	[one_star, one_star, one_star, one_star, one_star]
30	[bad, bad, bad, bad, good]	[one_star, one_star, one_star, one_star, one_star]
31	[bad, bad, bad, bad, bad]	[one_star, one_star, one_star, one_star, one_star]

These calculations are with $Pr(\texttt{five star}) = .3$, $Pr(\texttt{one star}) = .7$, $Pr(\texttt{good} \mid \texttt{five star}) = .95$, $Pr(\texttt{bad} \mid \texttt{five star}) = .05$, $Pr(\texttt{good} \mid \texttt{one star}) = .4$, $Pr(\texttt{bad} \mid \texttt{one star}) = .6$.

no.	state	algorithm
0	[good, good, good, good, good]	[five, five, five, five, five]
1	[good, good, good, good, bad]	[five, five, five, five, five]
2	[good, good, good, bad, good]	[five, five, five, one, five]
3	[good, good, good, bad, bad]	[five, five, five, one, one]
4	[good, good, bad, good, good]	[five, five, one, one, one]
5	[good, good, bad, good, bad]	[five, five, one, one, one]
6	[good, good, bad, bad, good]	[five, five, one, one, one]
7	[good, good, bad, bad, bad]	[five, five, one, one, one]
8	[good, bad, good, good, good]	[five, one, one, one, one]
9	[good, bad, good, good, bad]	[five, one, one, one, one]
10	[good, bad, good, bad, good]	[five, one, one, one, one]
11	[good, bad, good, bad, bad]	[five, one, one, one, one]
12	[good, bad, bad, good, good]	[five, one, one, one, one]
13	[good, bad, bad, good, bad]	[five, one, one, one, one]
14	[good, bad, bad, bad, good]	[five, one, one, one, one]
15	[good, bad, bad, bad, bad]	[five, one, one, one, one]

no.	state	algorithm
16	[bad, good, good, good, good]	[one, one, one, one, one]
17	[bad, good, good, good, bad]	[one, one, one, one, one]
18	[bad, good, good, bad, good]	[one, one, one, one, one]
19	[bad, good, good, bad, bad]	[one, one, one, one, one]
20	[bad, good, bad, good, good]	[one, one, one, one, one]
21	[bad, good, bad, good, bad]	[one, one, one, one, one]
22	[bad, good, bad, bad, good]	[one, one, one, one, one]
23	[bad, good, bad, bad, bad]	[one, one, one, one, one]
24	[bad, bad, good, good, good]	[one, one, one, one, one]
25	[bad, bad, good, good, bad]	[one, one, one, one, one]
26	[bad, bad, good, bad, good]	[one, one, one, one, one]
27	[bad, bad, good, bad, bad]	[one, one, one, one, one]
28	[bad, bad, bad, good, good]	[one, one, one, one, one]
29	[bad, bad, bad, good, bad]	[one, one, one, one, one]
30	[bad, bad, bad, bad, good]	[one, one, one, one, one]
31	[bad, bad, bad, bad, bad]	[one, one, one, one, one]

You can see that a cascade is predicted, but in a different way than before.

Note especially the recovery for the five-star vote at the end of number 2.