Reasoning about Knowledge

Thomas Ågotnes

University of Bergen, Norway

Examples

麻将

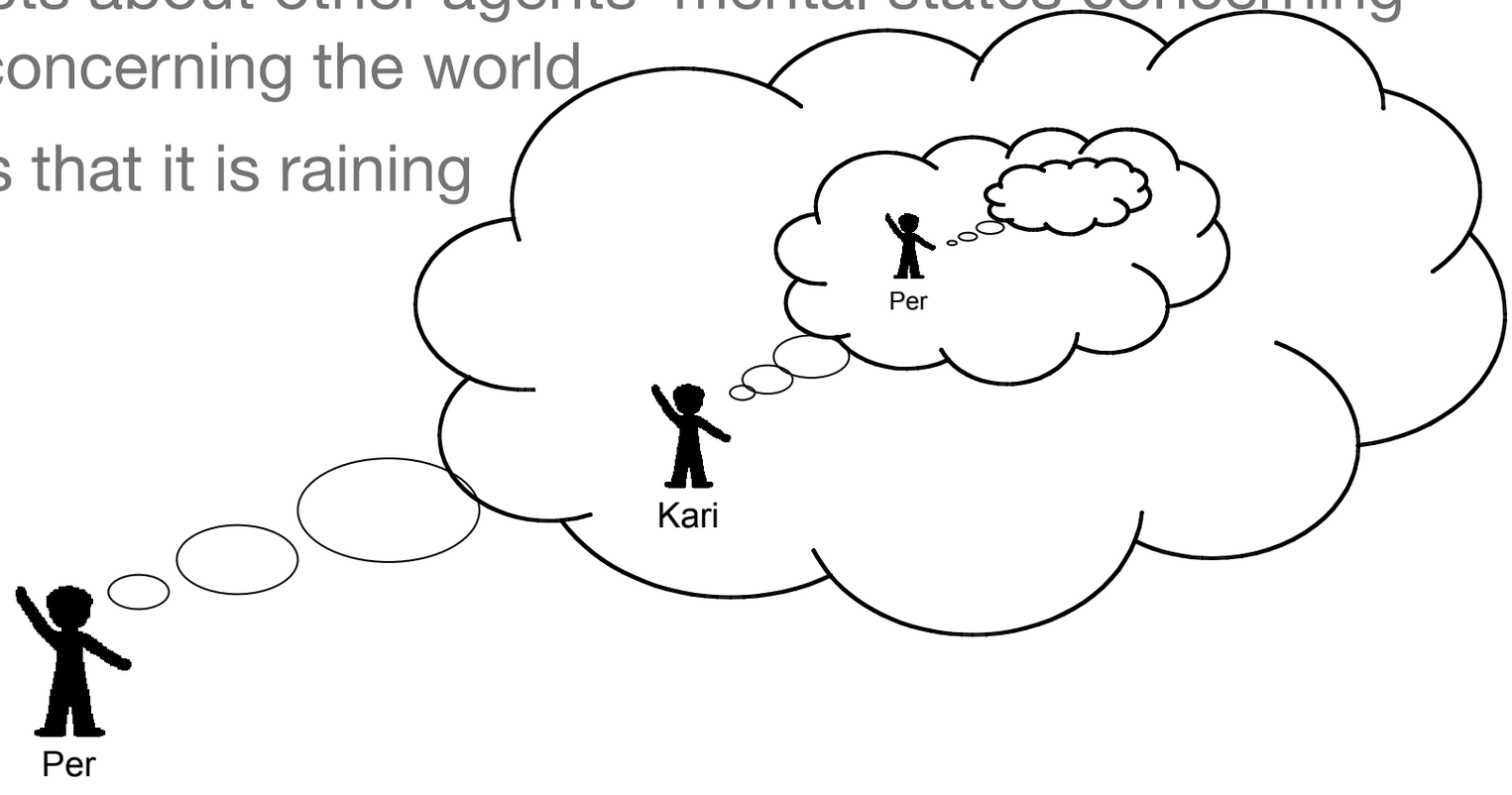
Card games

Example (American newspaper)

“Dean doesn't know whether Nixon knows that Dean knows that Nixon knows that McCord burgled O'Brien's office at Watergate”

Natural intelligence: higher-order social cognition

- Theory of mind: cognitive capacity to understand/predict external behaviour of others/oneself by attributing internal mental states
- 0th order TOM: concerns facts about the world
- 1st order TOM: concerns facts about other agents' mental states concerning the world
  - Bill knows that it is raining
- 2nd order TOM: concerns facts about other agents' mental states concerning other agents' mental states concerning the world
  - Ann knows that Bill knows that it is raining
- ...



Natural intelligence: higher-order social cognition

- Animals: probably no higher-order cognition
- Children: first-order reasoning
- Adults: passable shot at second-order reasoning
- ...

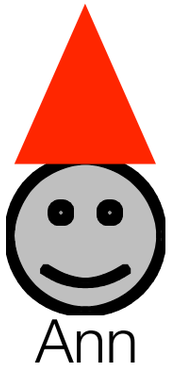
Today..

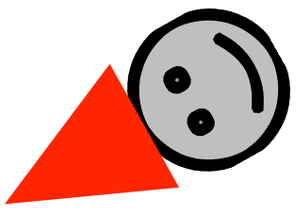
- .. I will only discuss *ideal* reasoning about knowledge
- .. which is not necessarily how humans always actually reason
- But it is a useful, and necessary, foundation, in order to understand the involved concepts
- We will define precisely what “knowledge” means, and how we can reason about it. And show that these definitions are useful in practice.

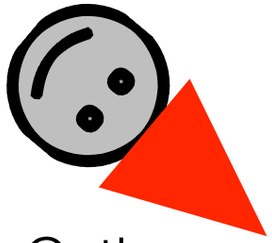
Reasoning about knowledge

- Epistemology: the study of knowledge
    - Ancient Greece
- Formal logical analysis
    - E.g., what are the axioms of knowledge?
- von Wright, 1950s
- Hintikka, 1962
- Recent decades: great interest from artificial intelligence, cognitive science, computer science, economics, linguistics, ...

3 Chinese students

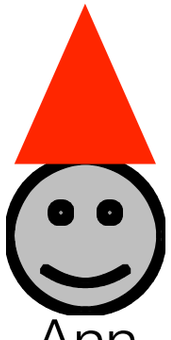




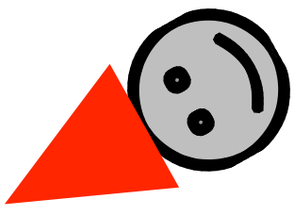


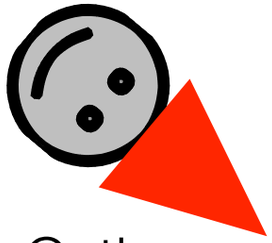
- All the hats are red
- You can only see the others' hats, not your own

3 Chinese students


Ann

- All the hats are red
- You can only see the others' hats, not your own

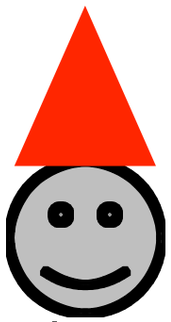

Bill


Cath

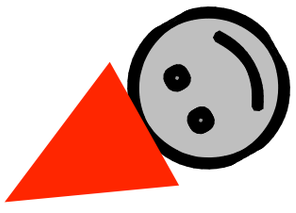
It is well known that Chinese students are:

- extremely intelligent
- honest
- perceptive

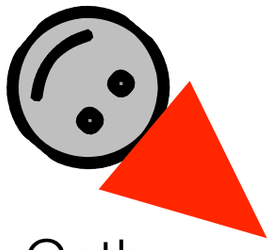
3 Chinese students



Ann



Bill



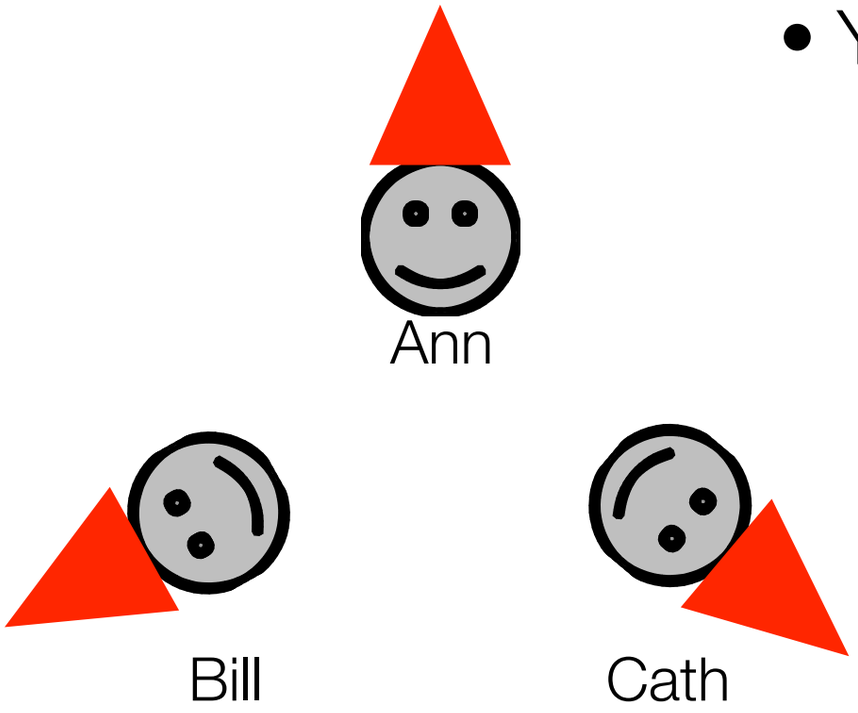
Cath

- All the hats are red
- You can only see the others' hats, not your own

teacher

3 Chinese students

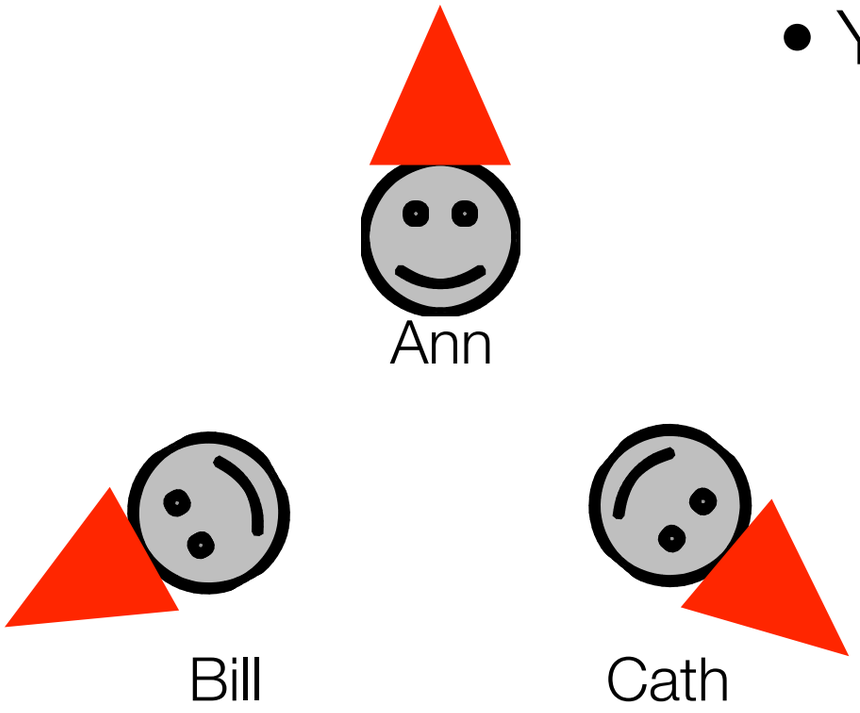
- All the hats are red
- You can only see the others' hats, not your own

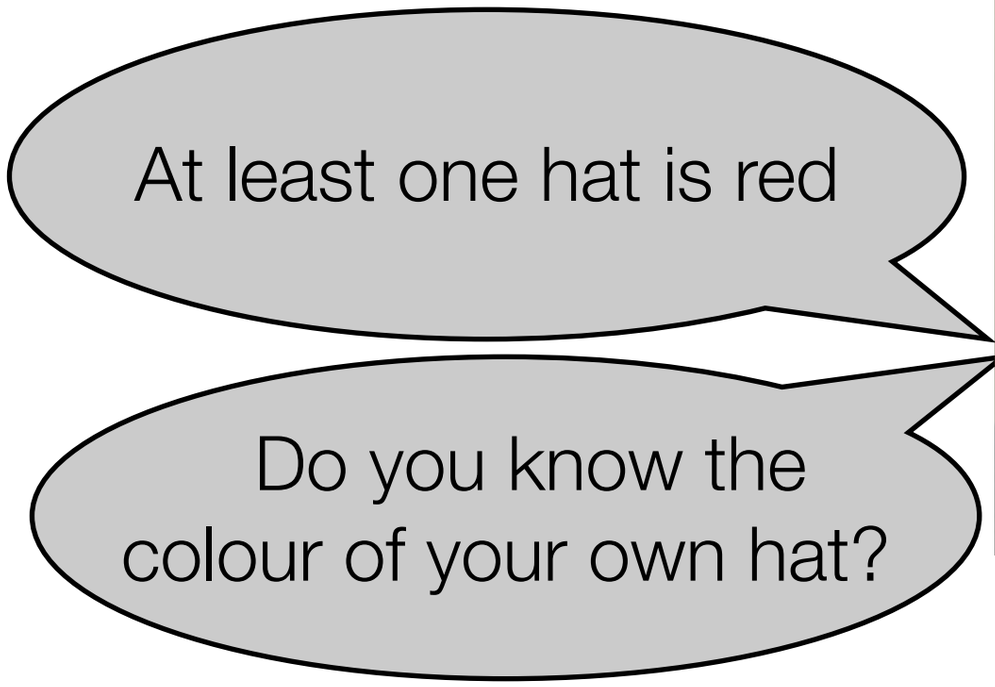


teacher

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own

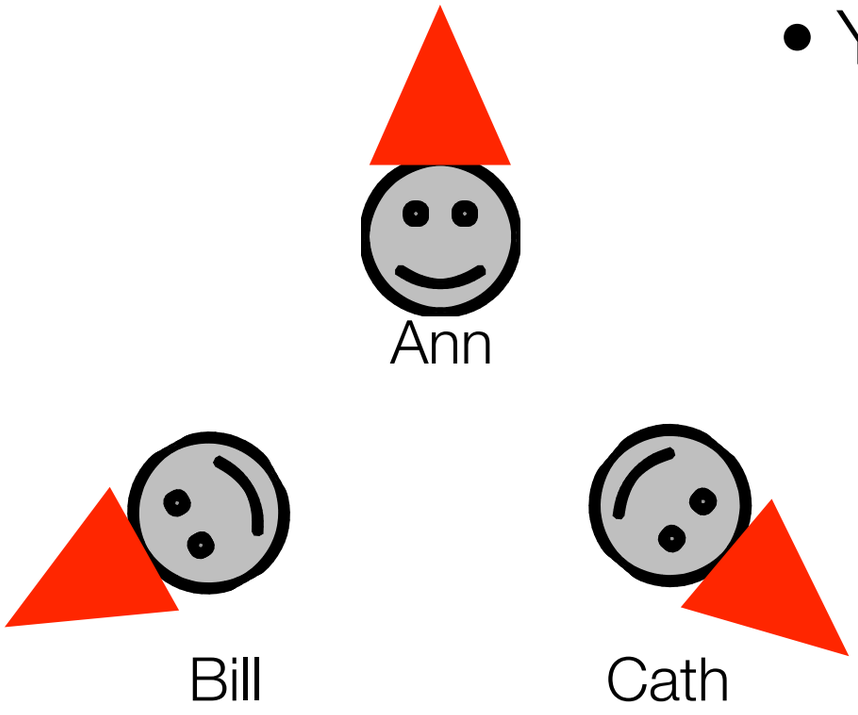


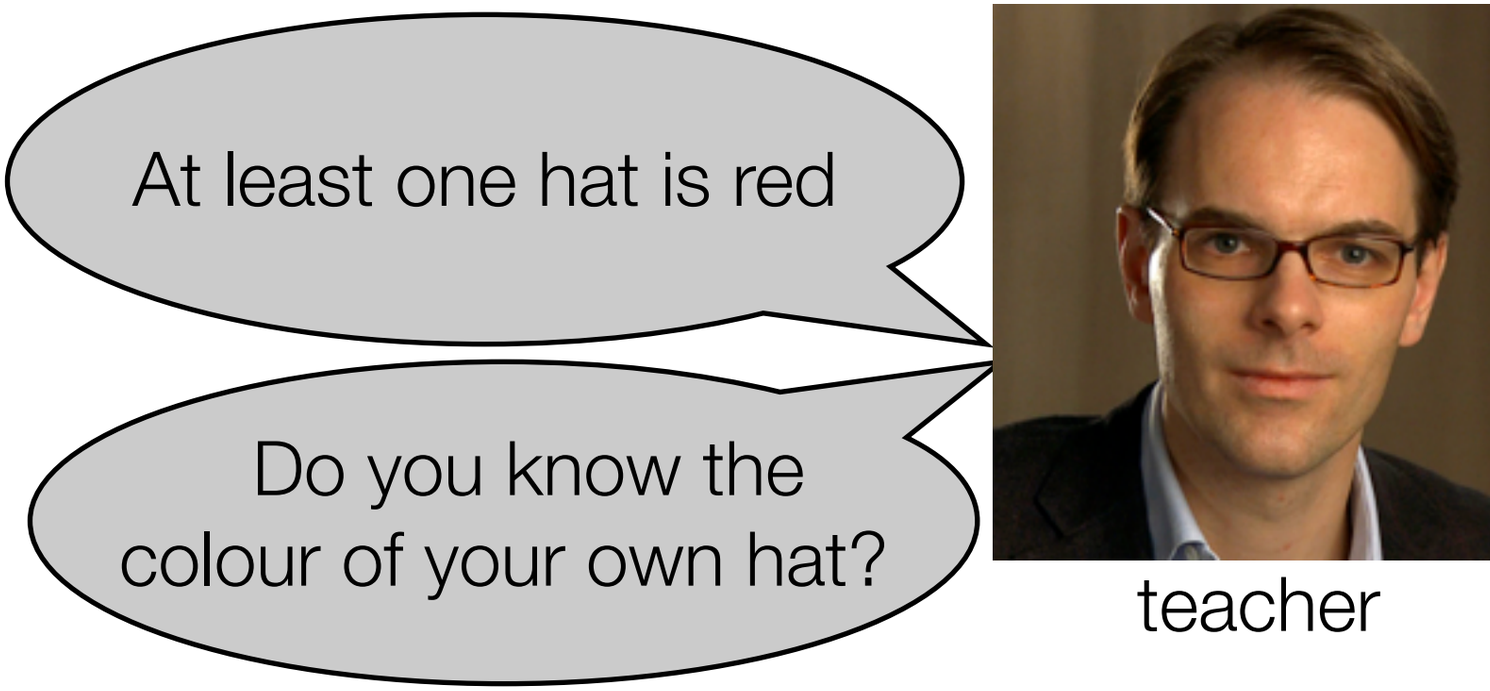


teacher

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own

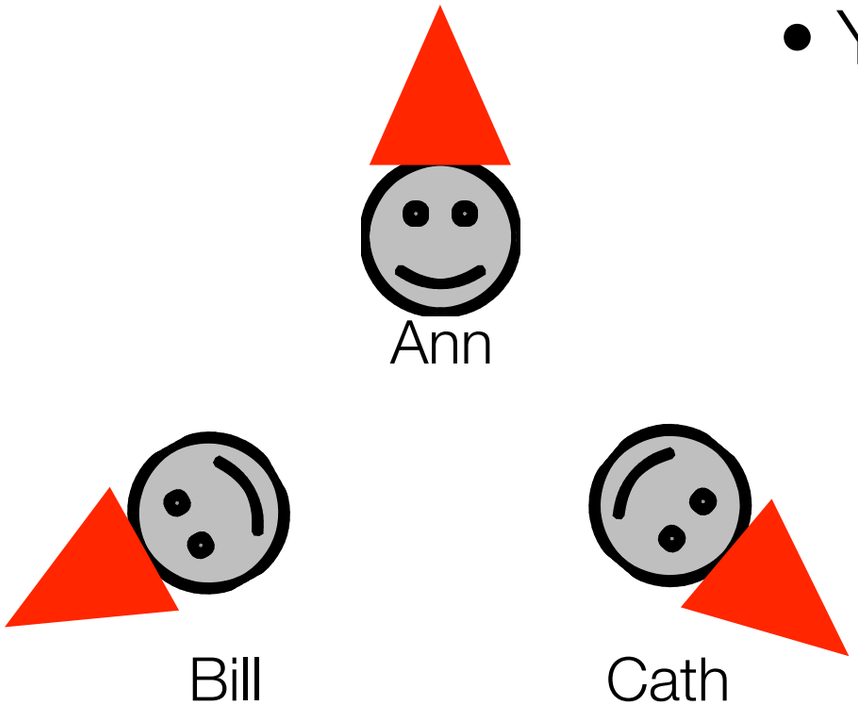


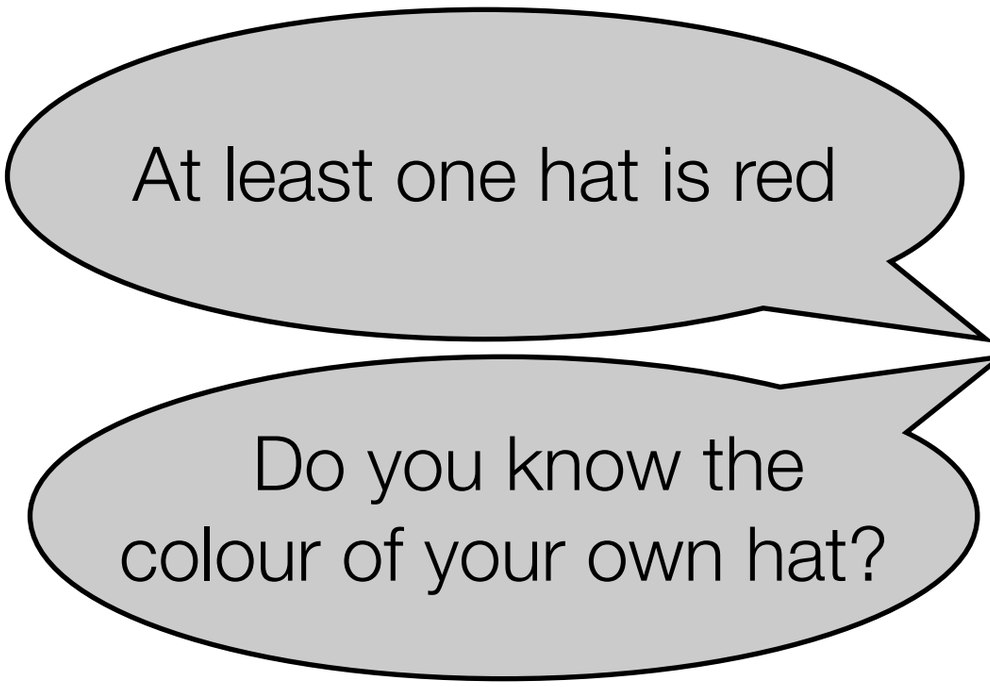


- All answer at the same time
- **What do they answer?**

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





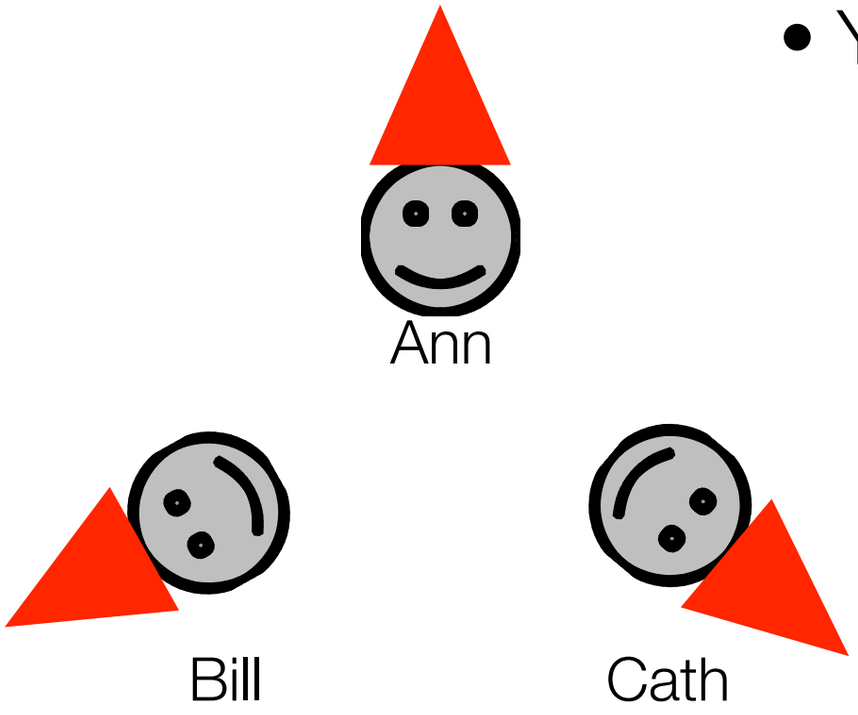

teacher

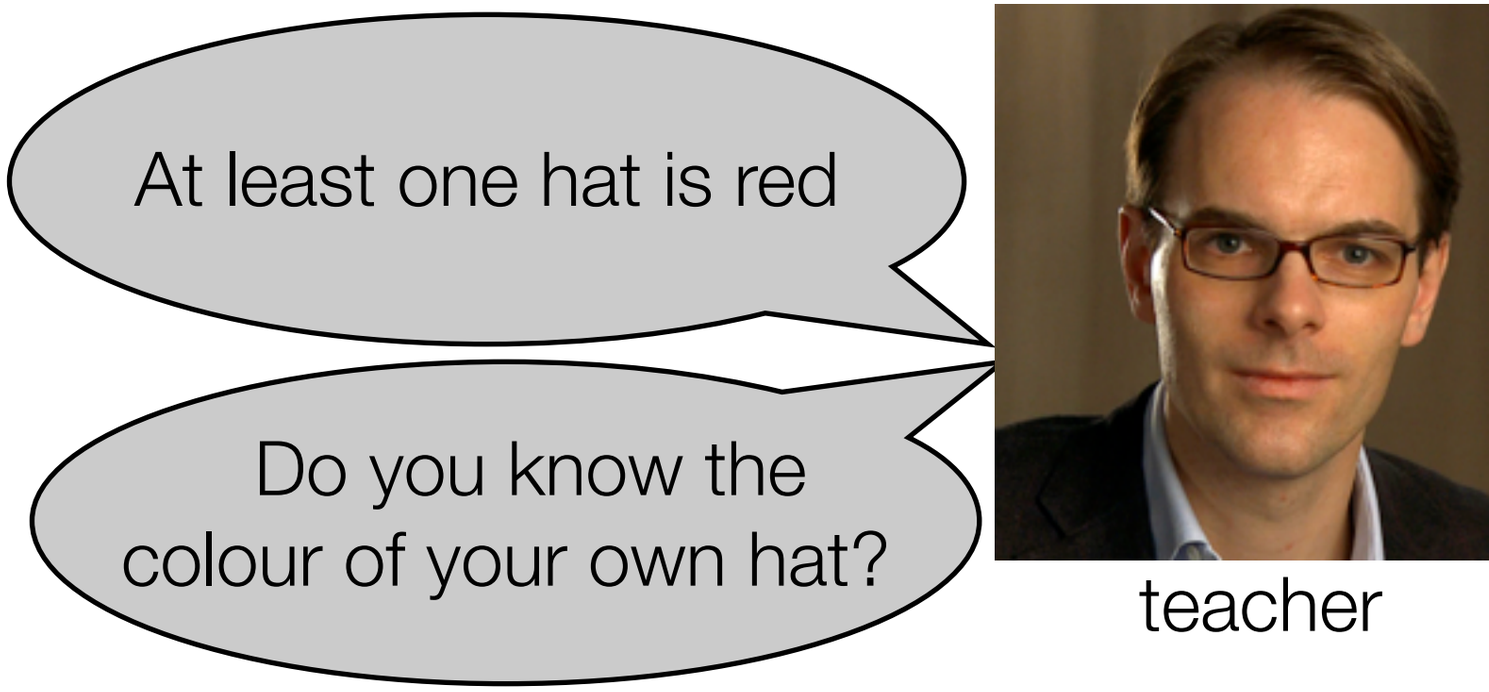
- All answer at the same time
- **What do they answer?**

- The teacher asks *again*
- What do they answer?
- etc.

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





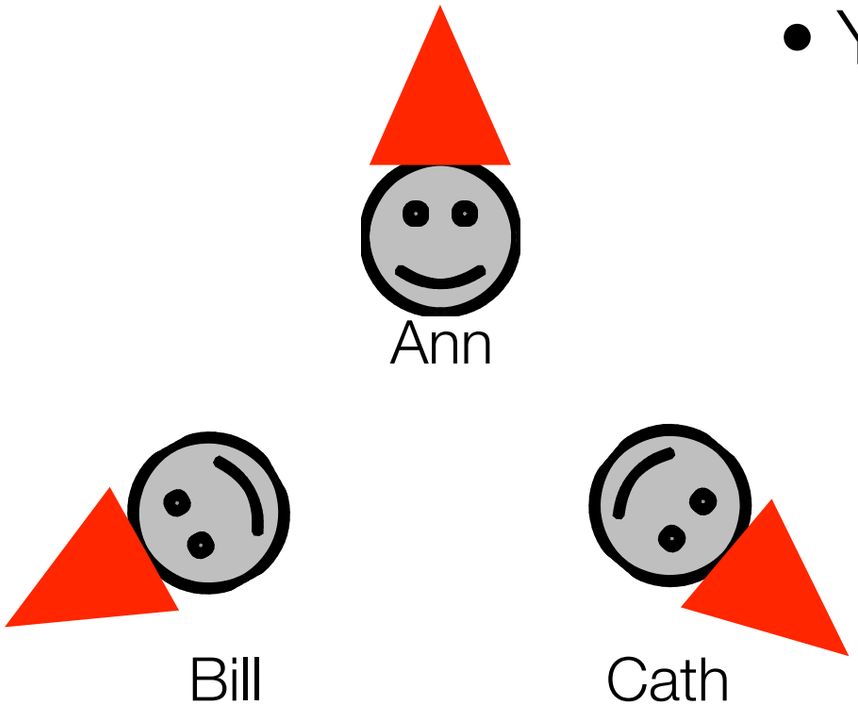
- All answer at the same time
- **What do they answer?**

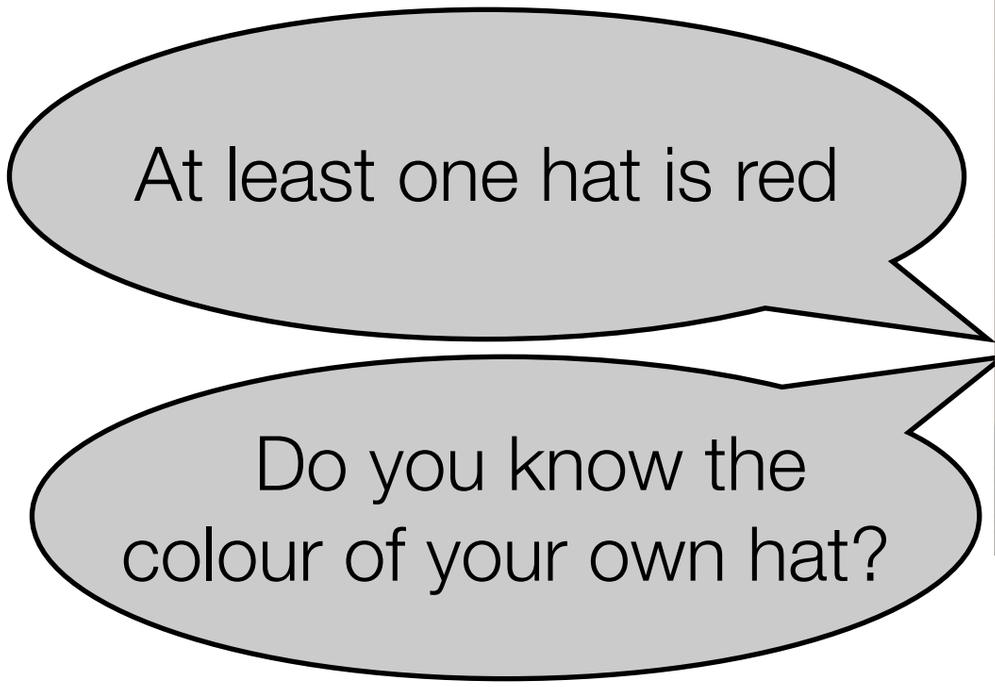
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own






teacher

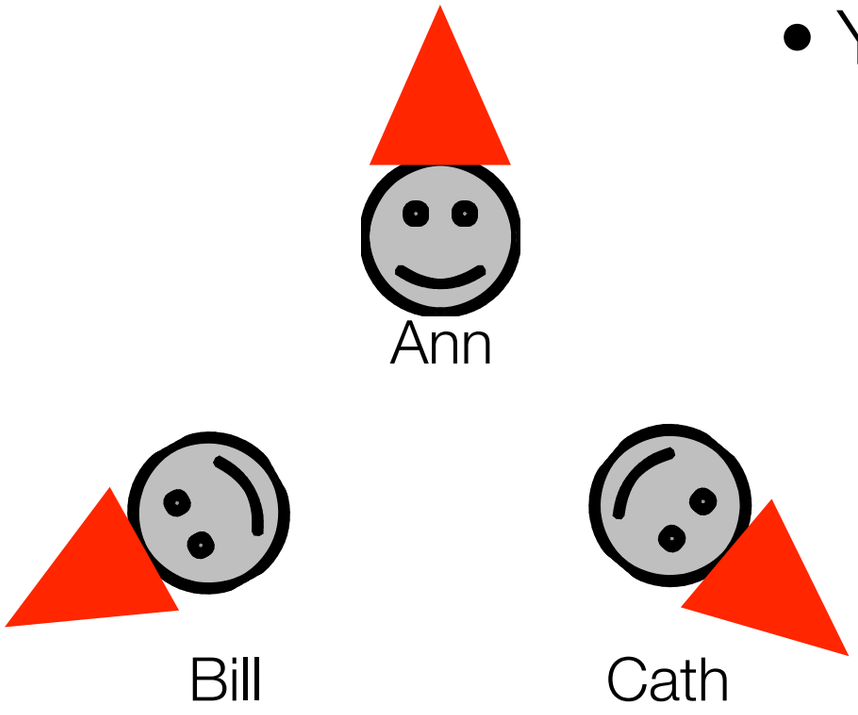
- All answer at the same time
- **What do they answer?**

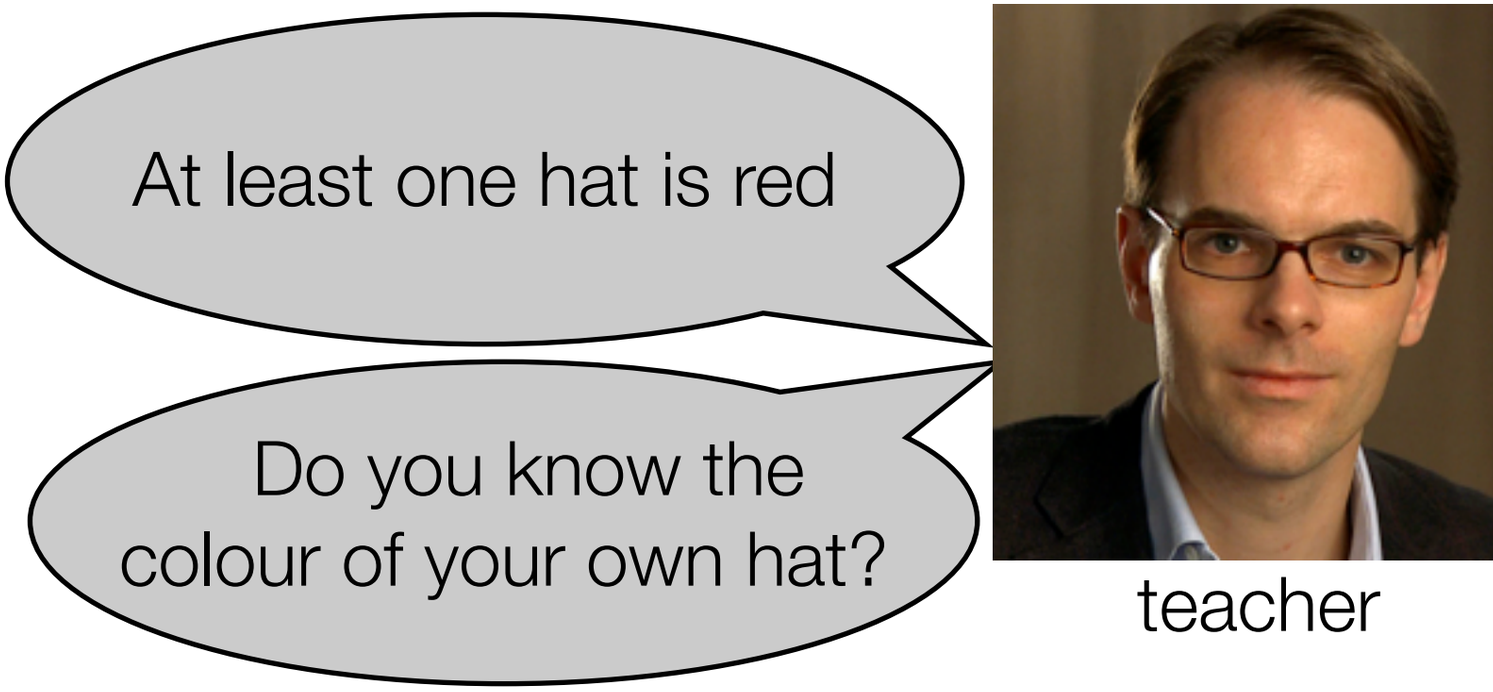
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1."no"

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





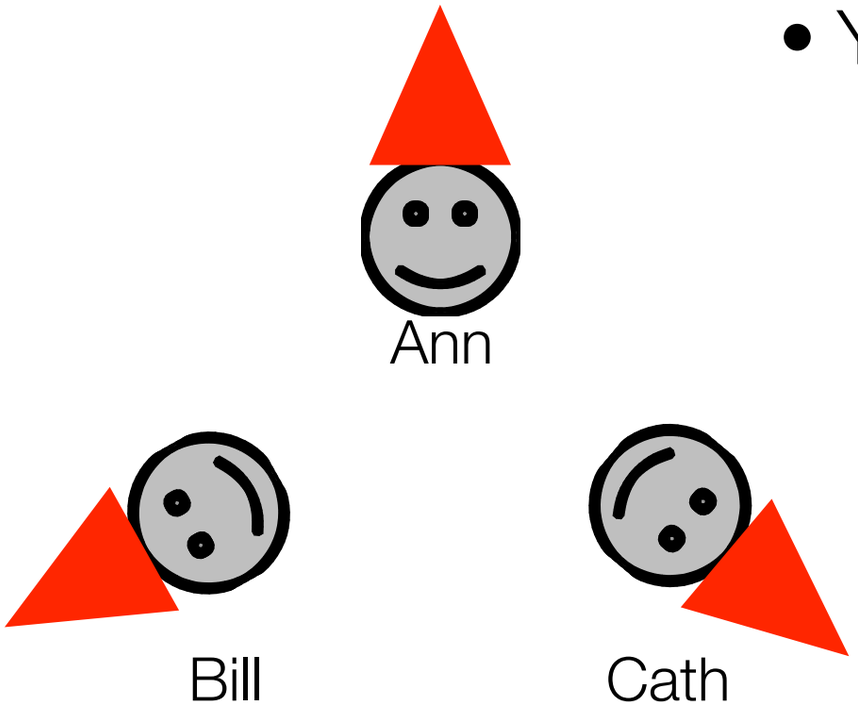
- All answer at the same time
- **What do they answer?**

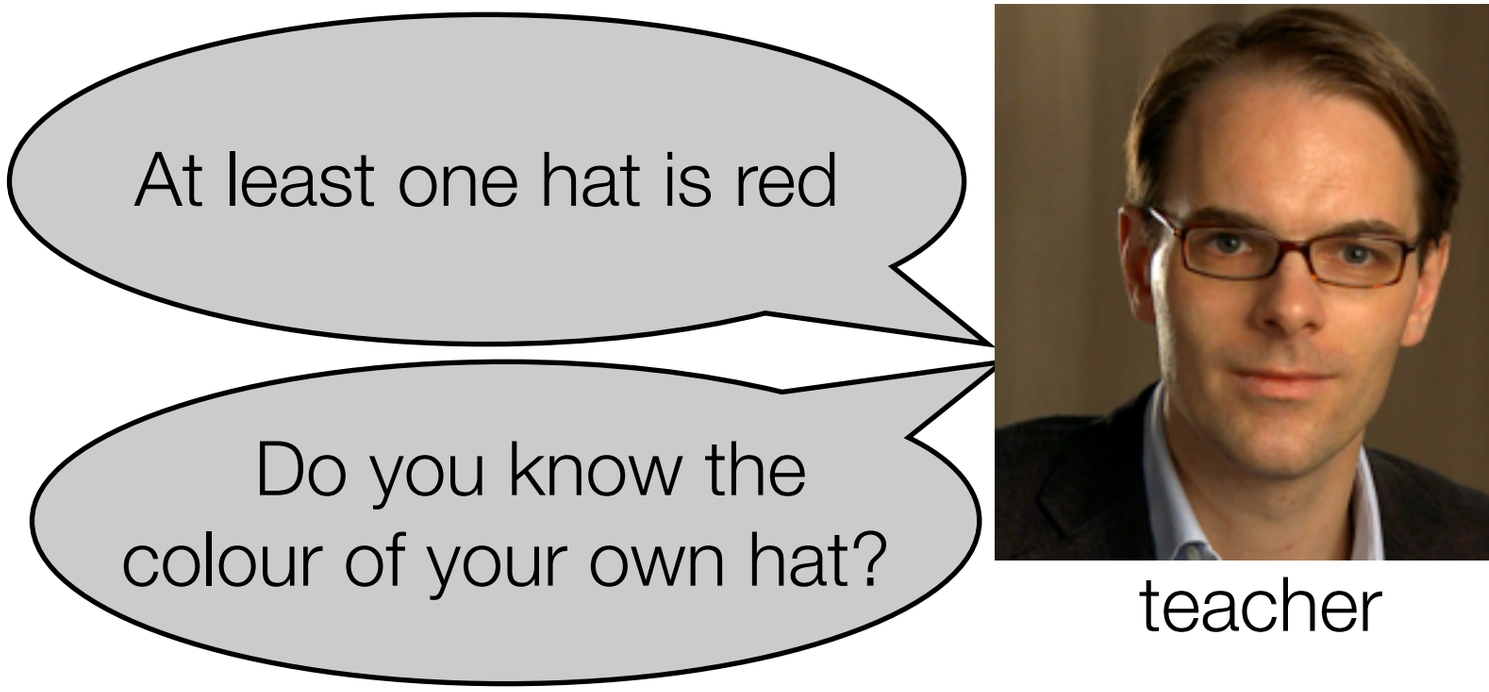
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1."no"
2."no"

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





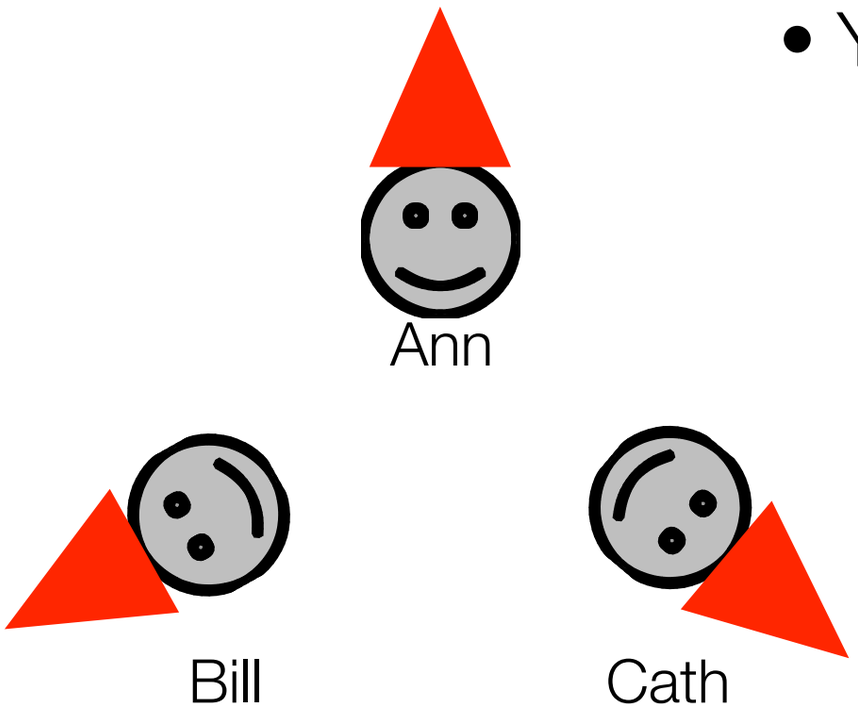
- All answer at the same time
- **What do they answer?**

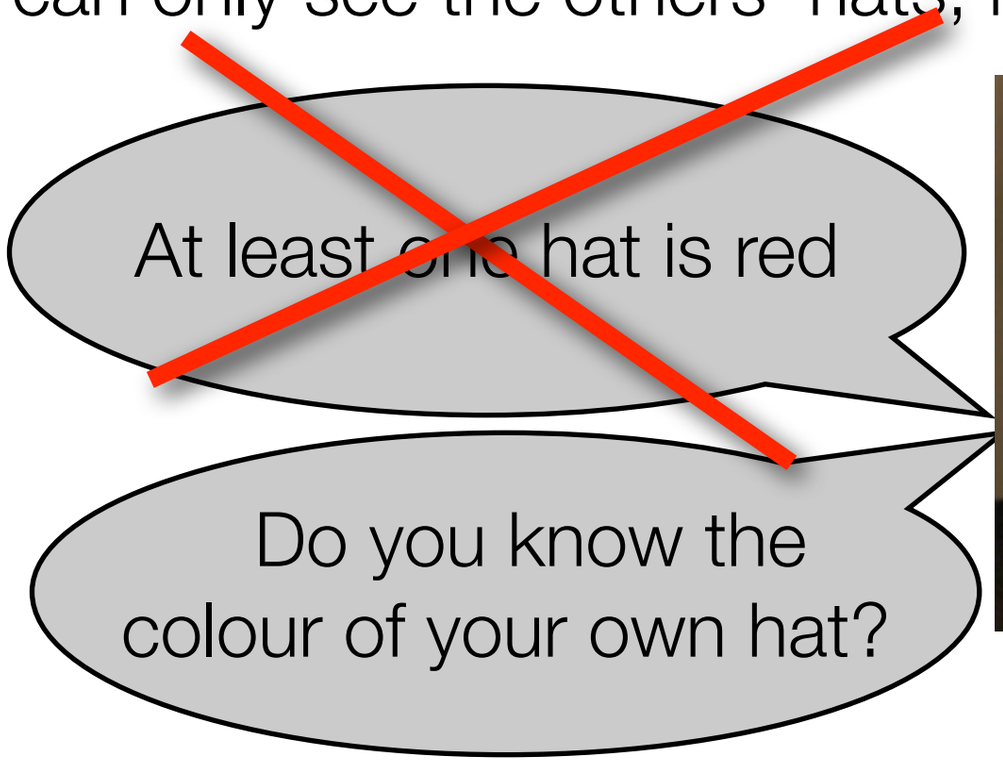
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1."no"
2."no"
3."yes"

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own



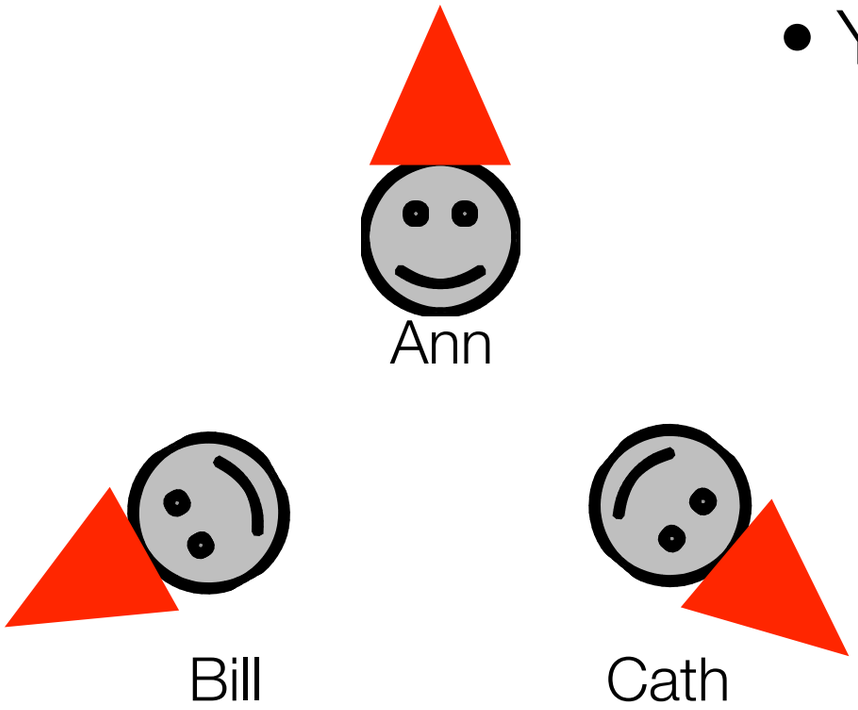


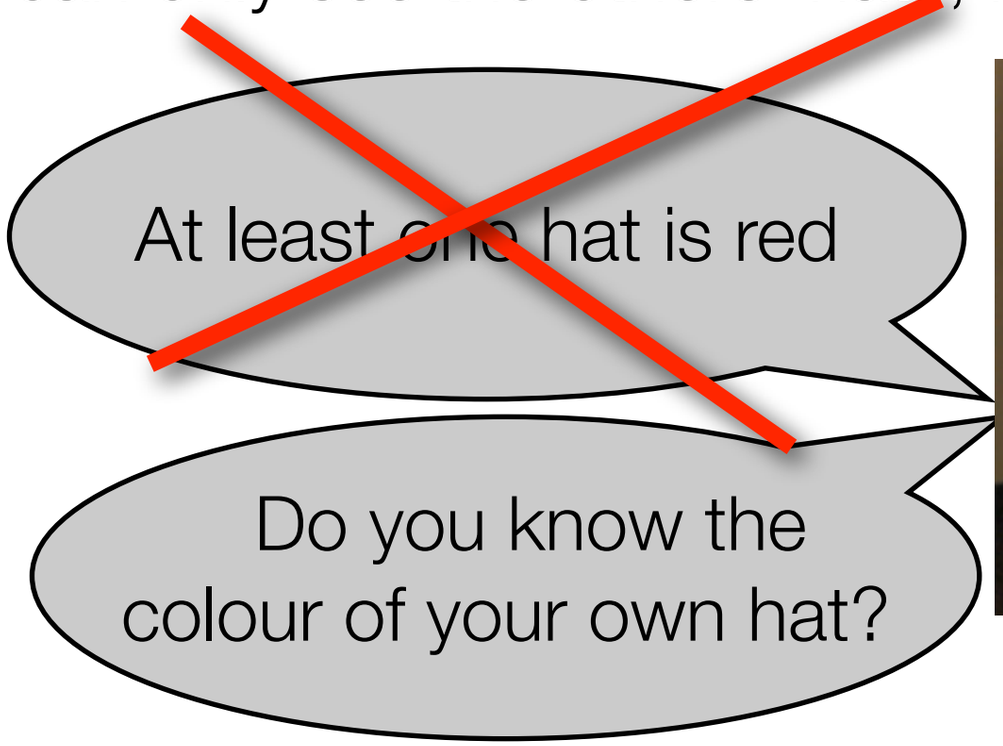
- All answer at the same time
- **What do they answer?**

- The teacher asks *again*
- What do they answer?
- etc.

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





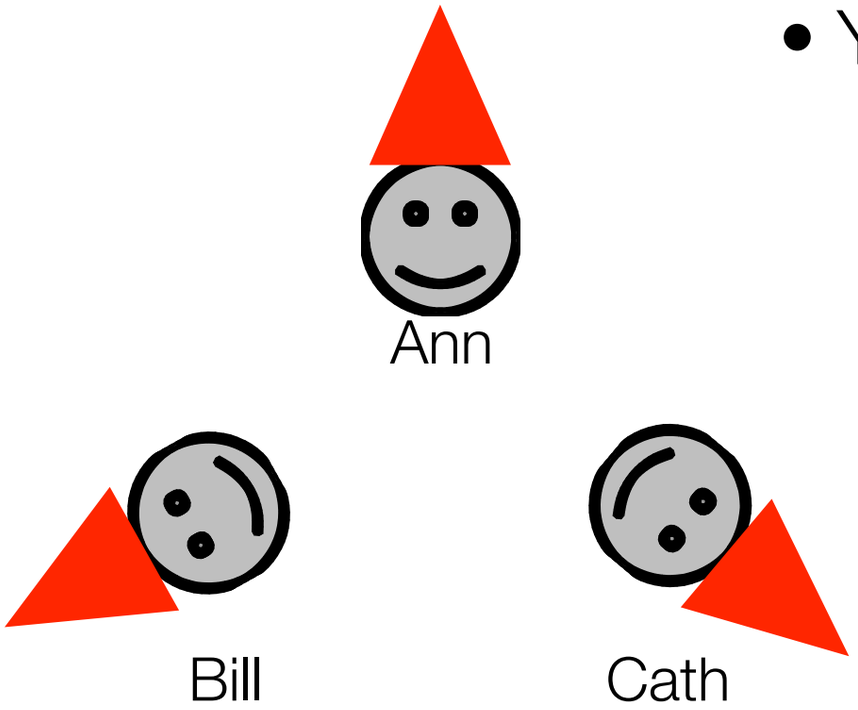
- All answer at the same time
- **What do they answer?**

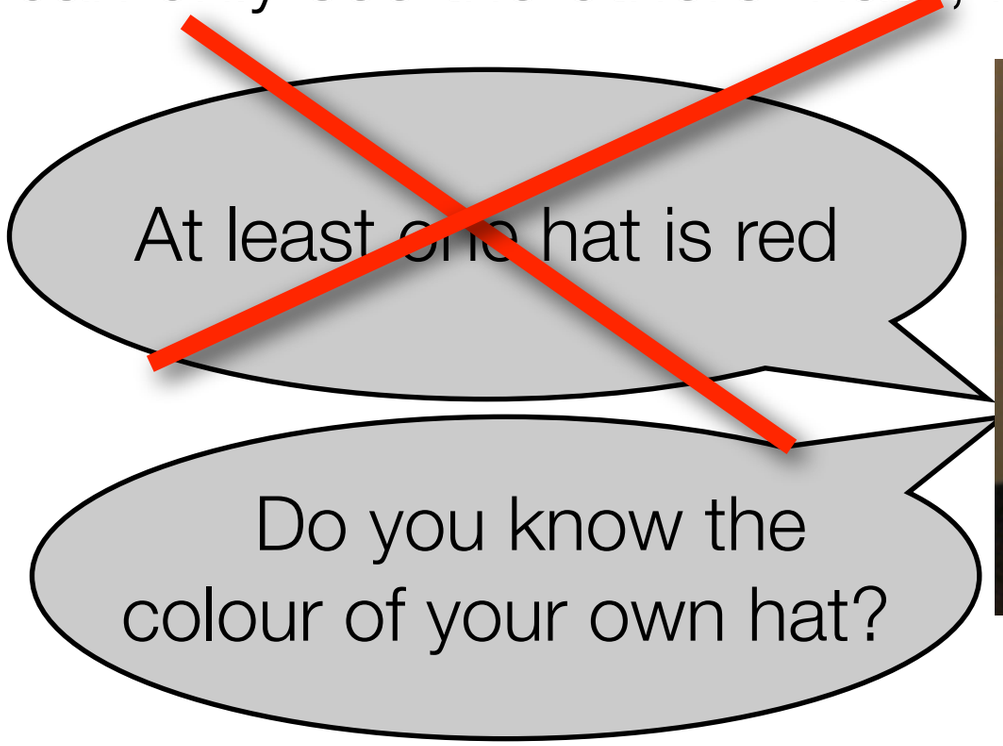
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





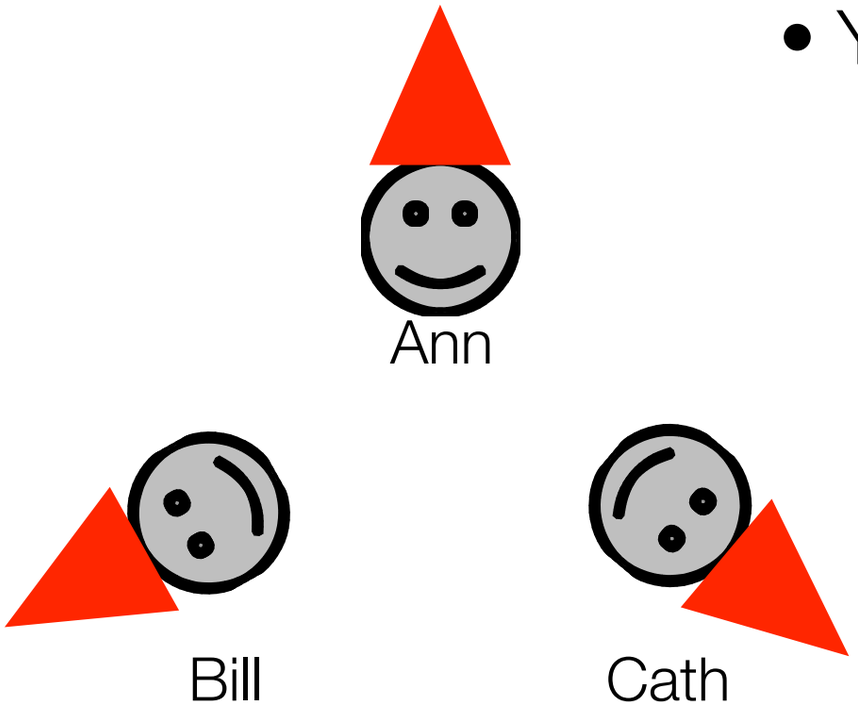
- All answer at the same time
- **What do they answer?**

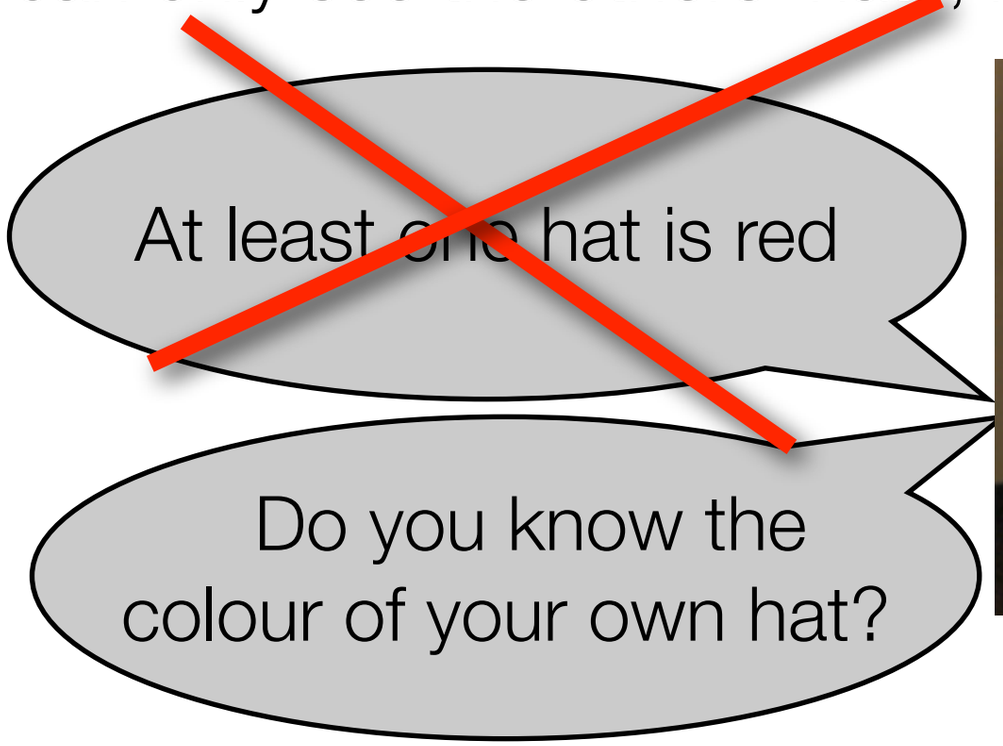
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1."no"

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





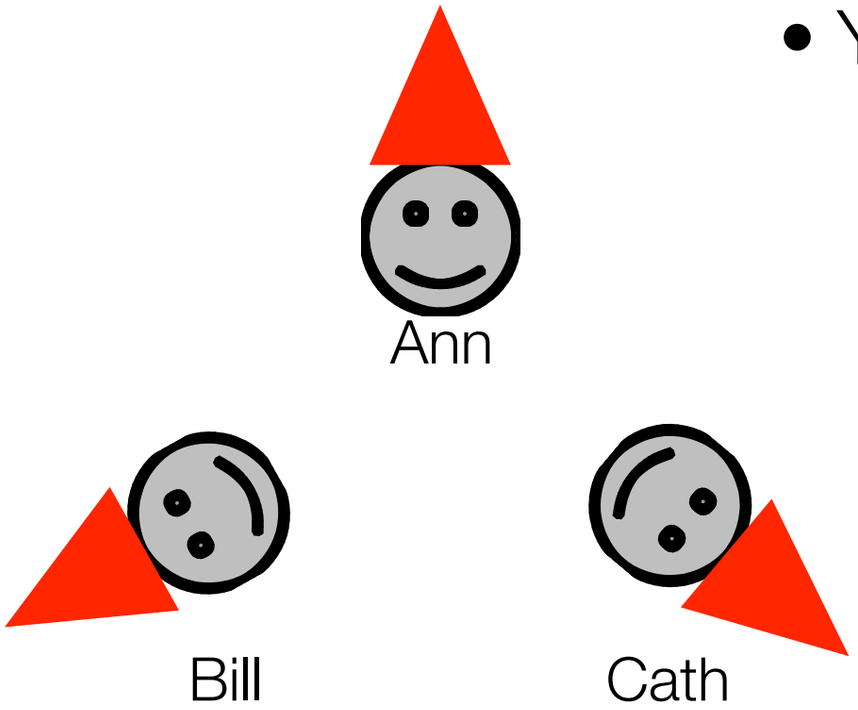
- All answer at the same time
- **What do they answer?**

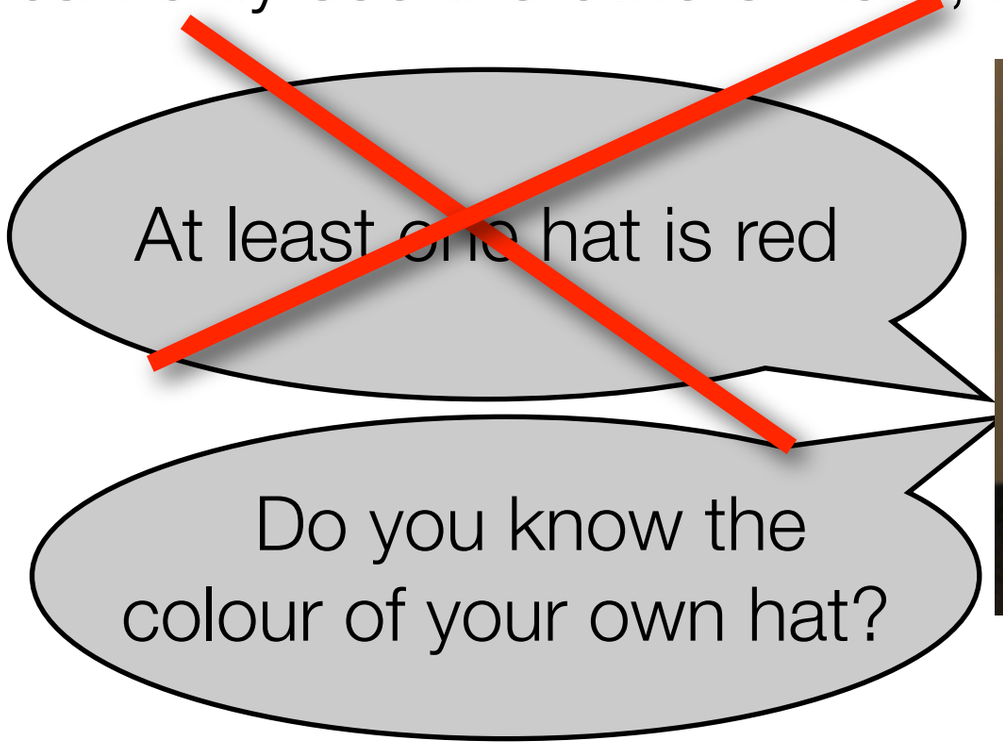
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1."no"
2."no"

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





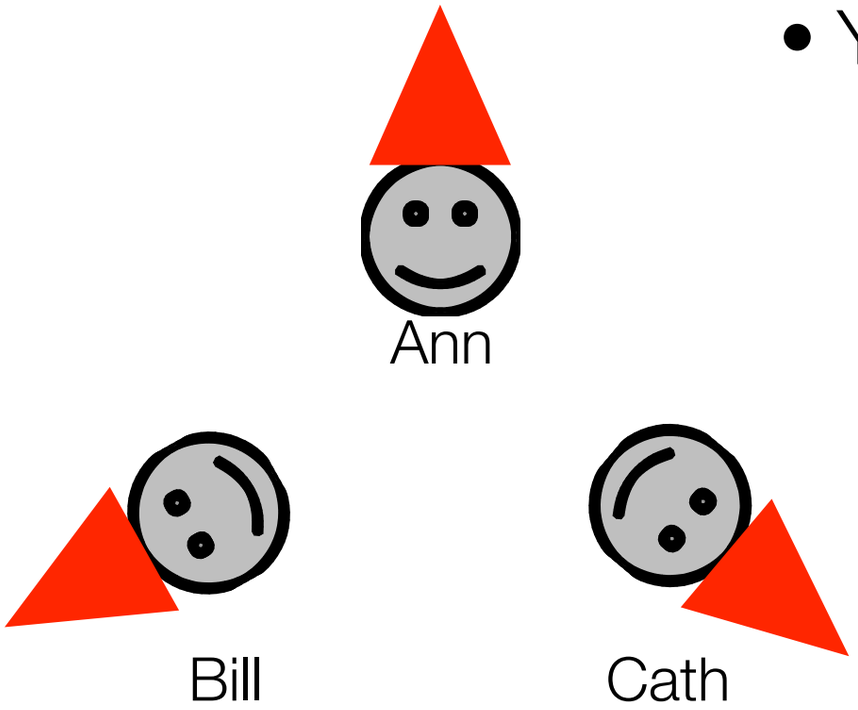
- All answer at the same time
- **What do they answer?**

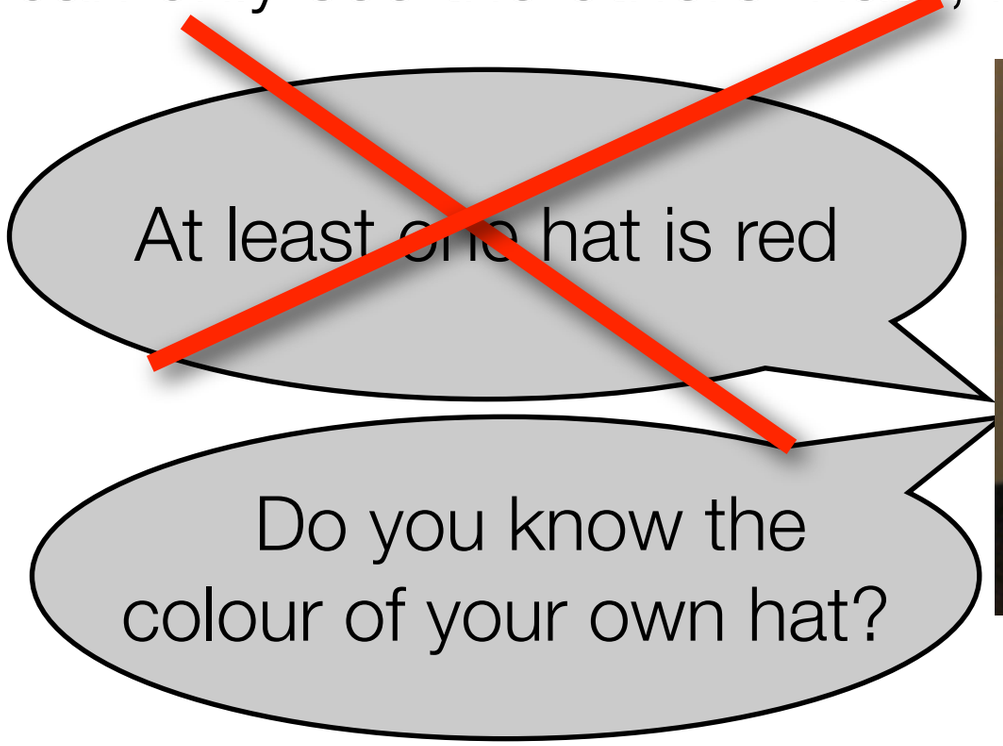
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1. "no"
2. "no"
3. "no"

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own





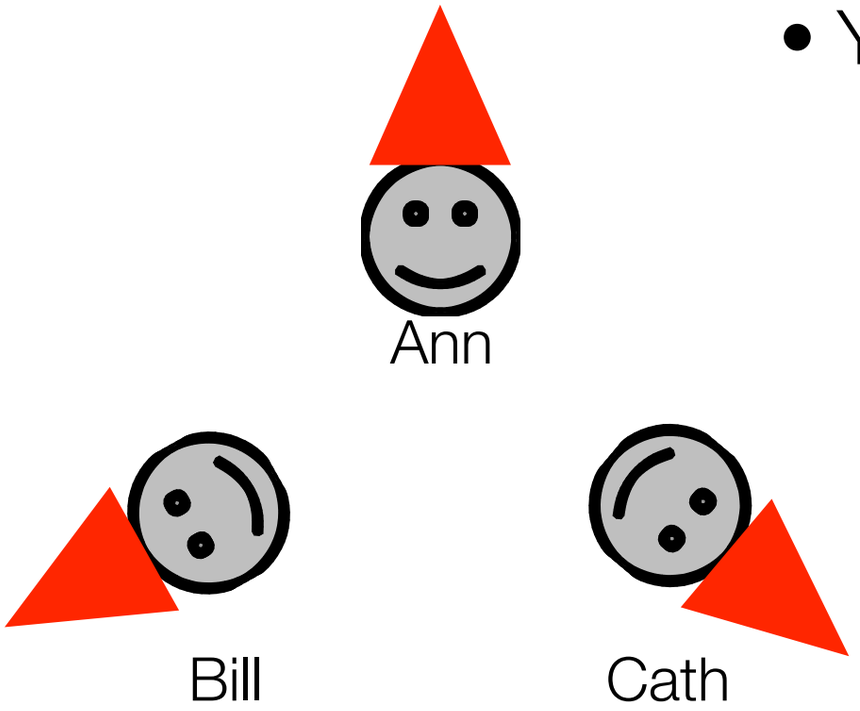
- All answer at the same time
- **What do they answer?**

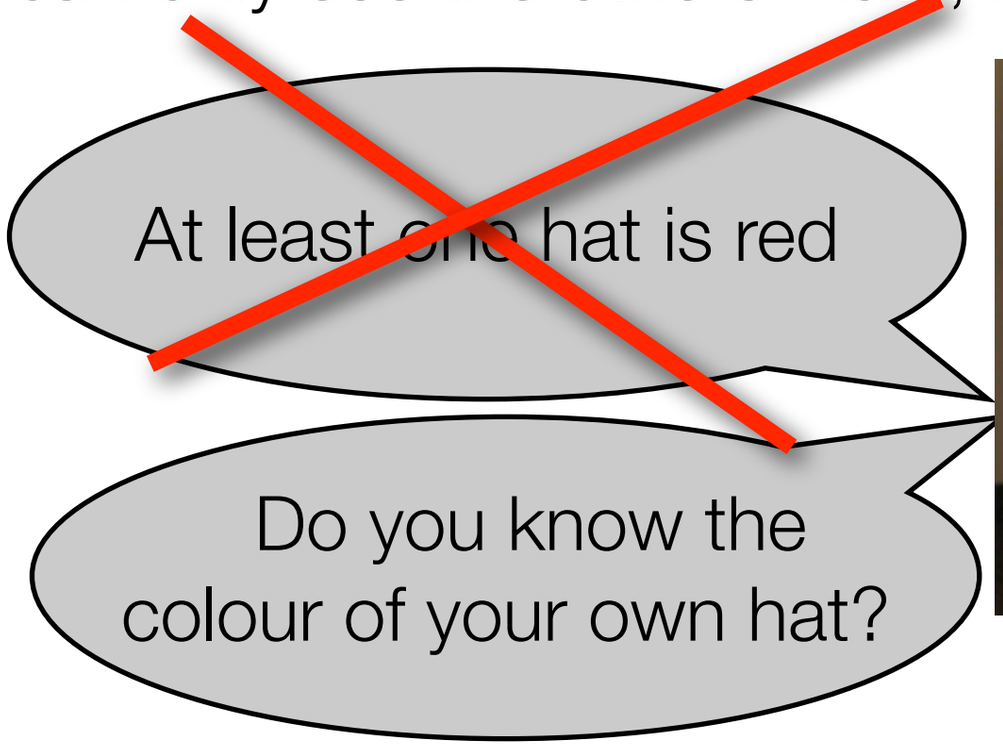
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1. "no"
2. "no"
3. "no"
4. "no"

3 Chinese students

- All the hats are red
- You can only see the others' hats, not your own



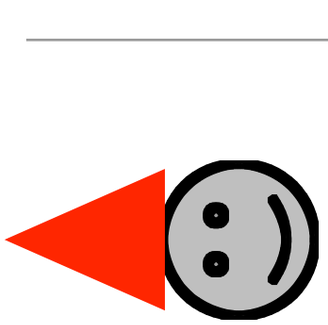


- All answer at the same time
- **What do they answer?**

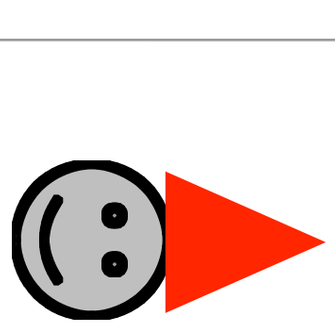
- The teacher asks *again*
- What do they answer?
- etc.

SOLUTION (all):
1. "no"
2. "no"
3. "no"
4. "no"
5. ...

Informal reasoning: simplified case of two students



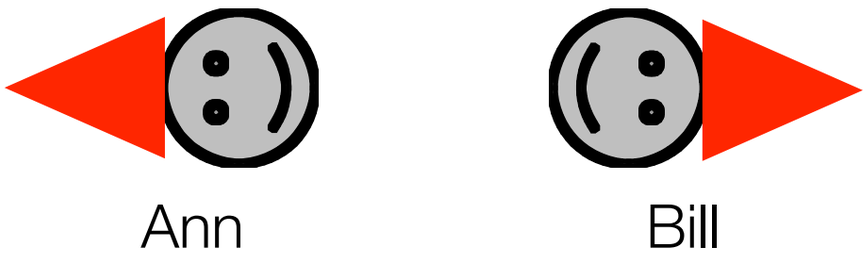
Ann



Bill

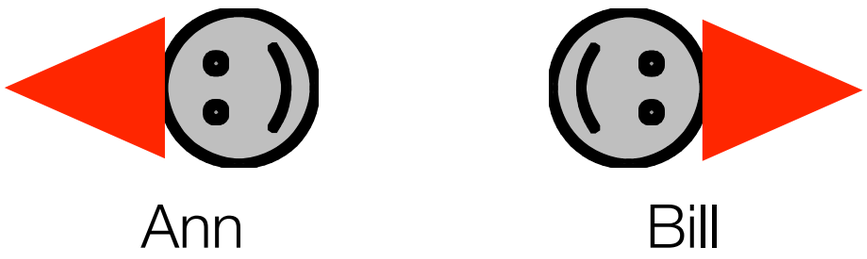
At least one hat is red

Informal reasoning: simplified case of two students



- How does Ann reason?

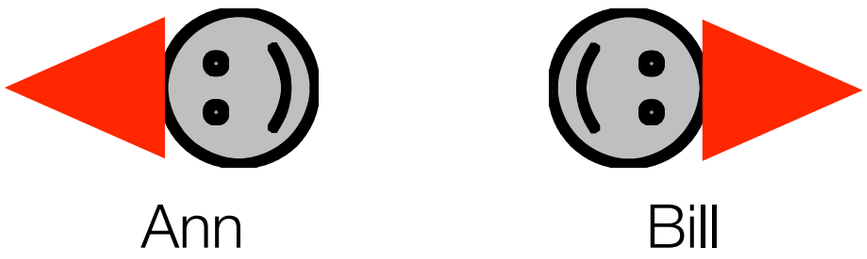
Informal reasoning: simplified case of two students



At least one hat is red

- How does Ann reason?
- After the teacher’s announcement, but before anyone answers:

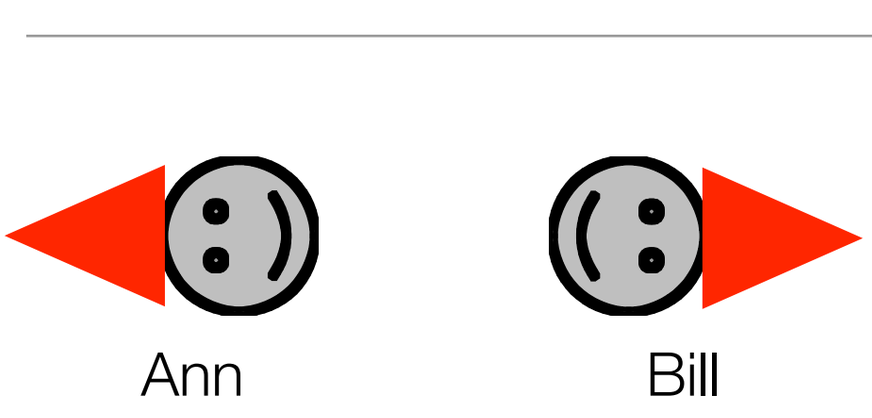
Informal reasoning: simplified case of two students



At least one hat is red

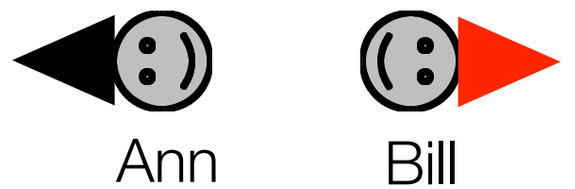
- How does Ann reason?
- After the teacher's announcement, but before anyone answers:
  - *Ann cannot discern between these situations*

Informal reasoning: simplified case of two students

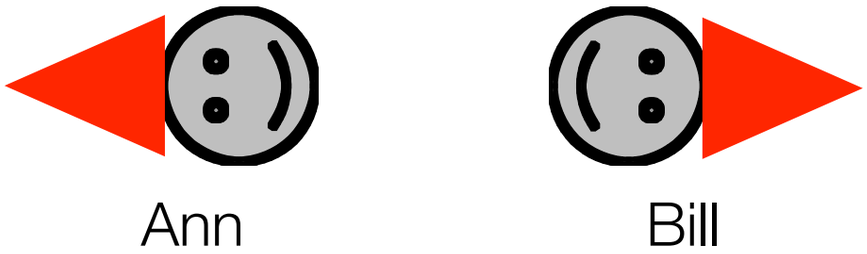


At least one hat is red

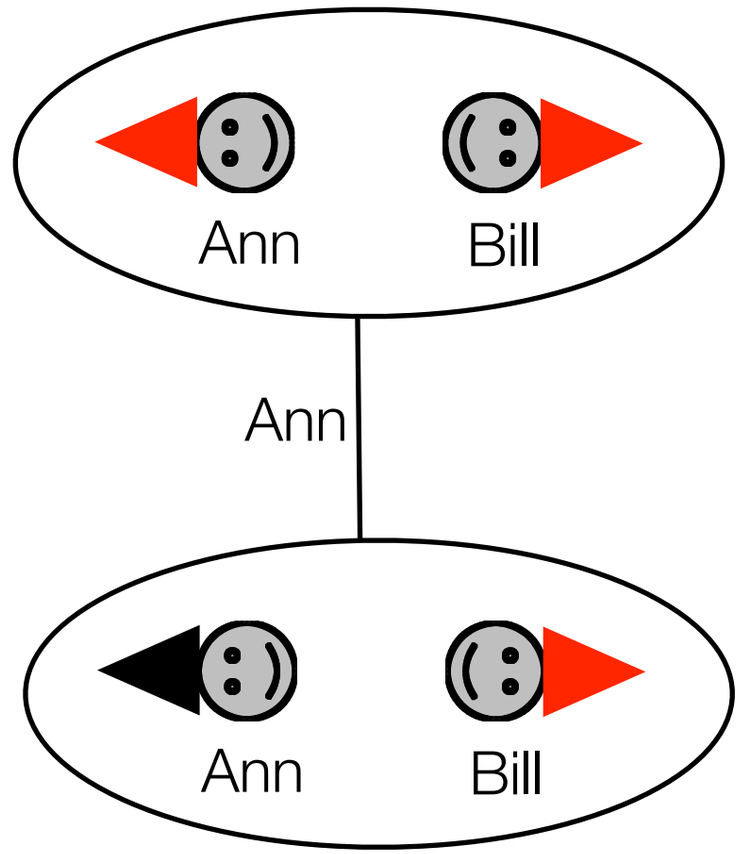
- How does Ann reason?
- After the teacher's announcement, but before anyone answers:
  - *Ann cannot discern between these situations*



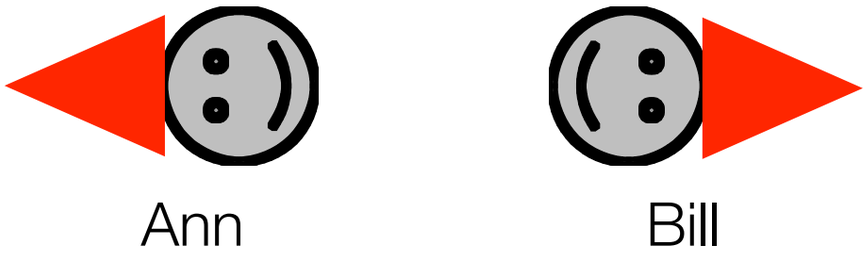
Informal reasoning: simplified case of two students



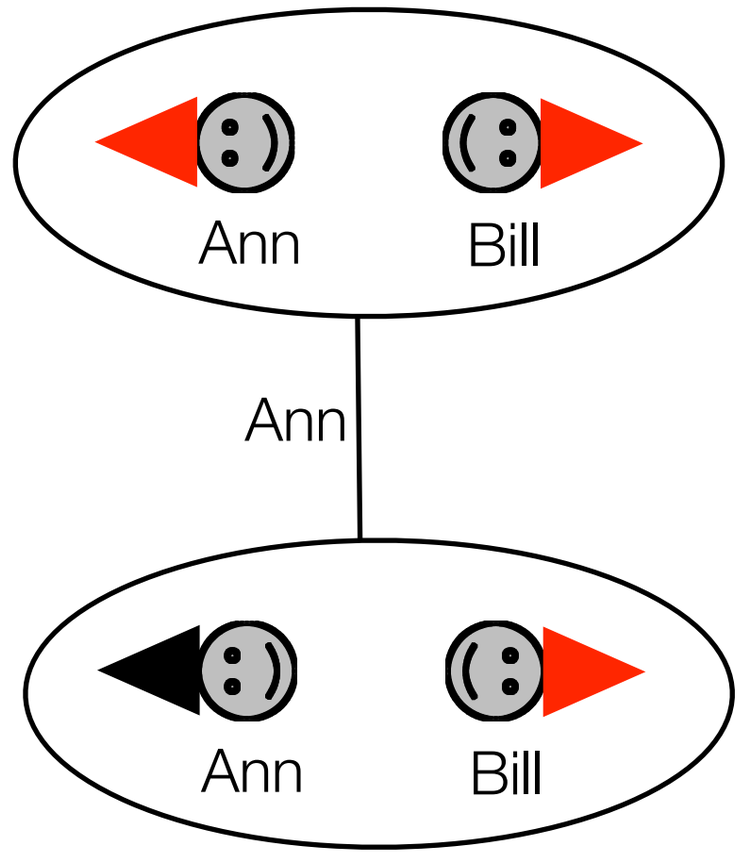
- How does Ann reason?
- After the teacher's announcement, but before anyone answers:
  - *Ann cannot discern between these situations*



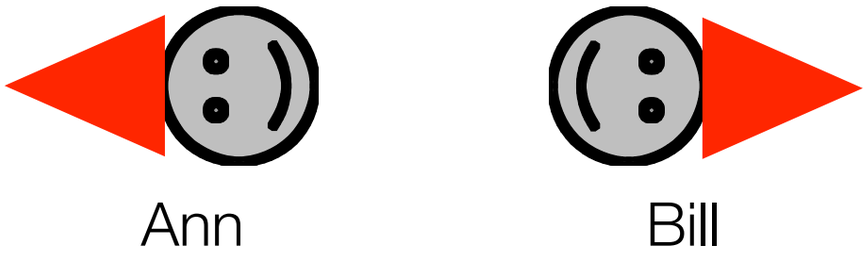
Informal reasoning: simplified case of two students



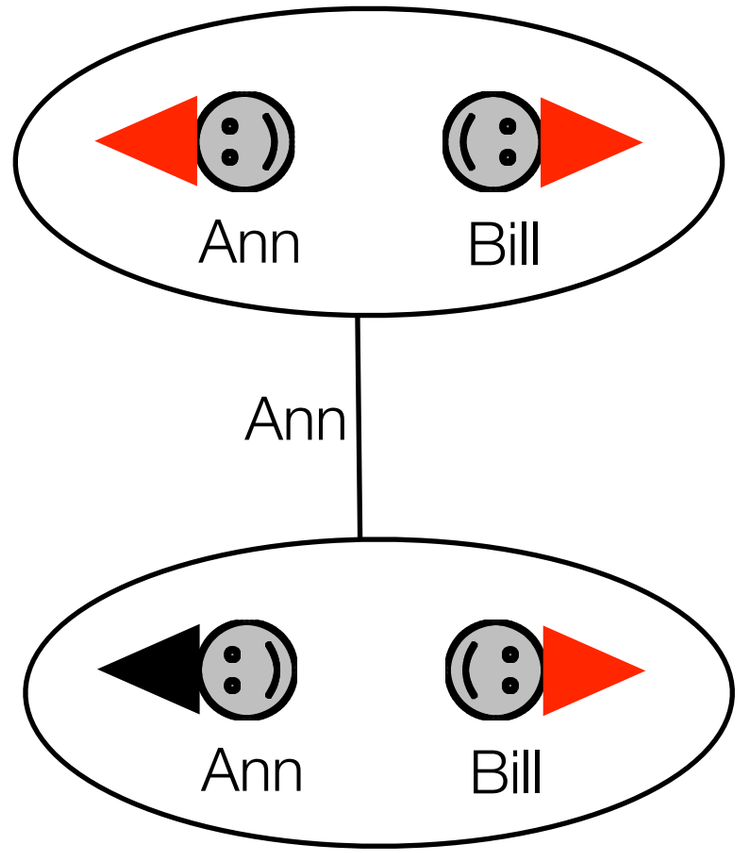
- How does Ann reason?
- After the teacher's announcement, but before anyone answers:
  - *Ann cannot discern between these situations*
  - She considers both situations *possible,* knows that *one of them* must be the *actual situation*, but does not know *which*



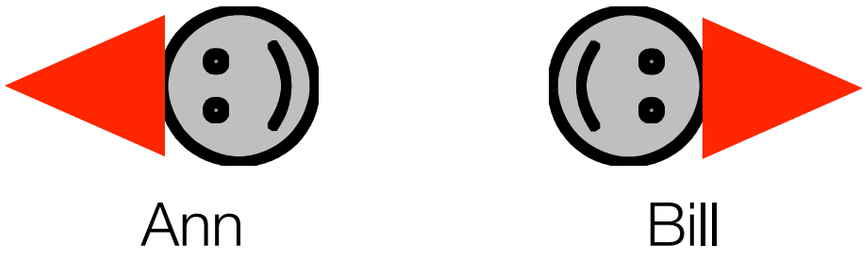
Informal reasoning: simplified case of two students



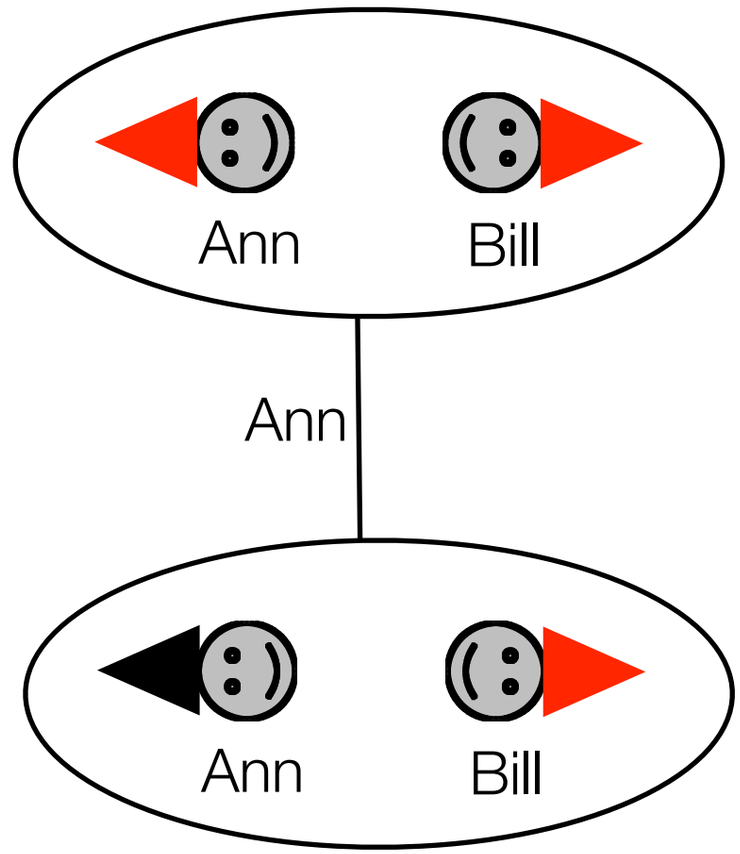
- How does Ann reason?
- After the teacher's announcement, but before anyone answers:
  - *Ann cannot discern between these situations*
  - She considers both situations *possible,* knows that *one of them* must be the *actual situation*, but does not know *which*
  - Thus, she answers "no"



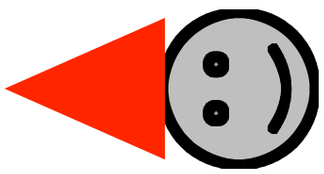
Informal reasoning: simplified case of two students



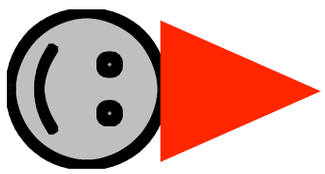
- How does Ann reason?
- After the teacher's announcement, but before anyone answers:
  - *Ann cannot discern between these situations*
  - She considers both situations *possible,* knows that *one of them* must be the *actual situation*, but does not know *which*
  - Thus, she answers "no"
  - And similarly for Bill



Informal reasoning: simplified case of two students



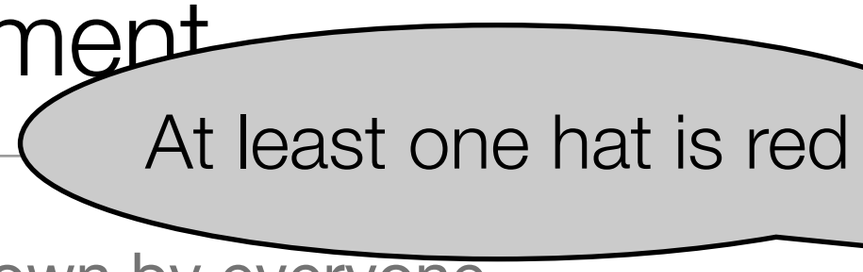
Ann



Bill

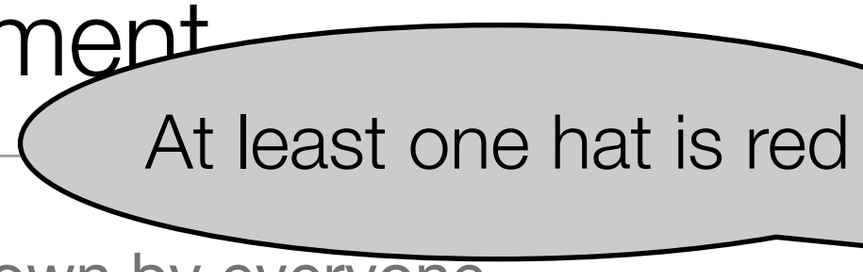
- How does Ann reason?
- After both Ann and Bill answered “no” to the first question:
  - Ann: “Bill knows that at least one hat is red, so if my hat is not red Bill would have known it (since he can see it) and would thus have known that his own hat must be red and answered “yes” to the first question. But he didn’t, so my hat must be red.
  - So Ann answers “yes” to the second question.
  - Similarly for Bill.

The role of the teacher's announcement



- The teacher announces something that is already known by everyone
- But which of the following statements are true (before the announcement)?
  - Ann knows that Bill knows that at least one hat is red
  - Bill knows that Ann knows that at least one hat is red
  - Ann knows that Bill knows that Cath knows that at least one hat is red

The role of the teacher's announcement



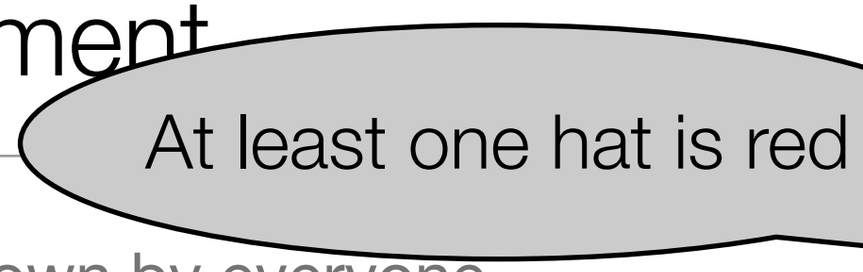
- The teacher announces something that is already known by everyone
- But which of the following statements are true (before the announcement)?
  - Ann knows that Bill knows that at least one hat is red YES
  - Bill knows that Ann knows that at least one hat is red
  - Ann knows that Bill knows that Cath knows that at least one hat is red

The role of the teacher's announcement

At least one hat is red

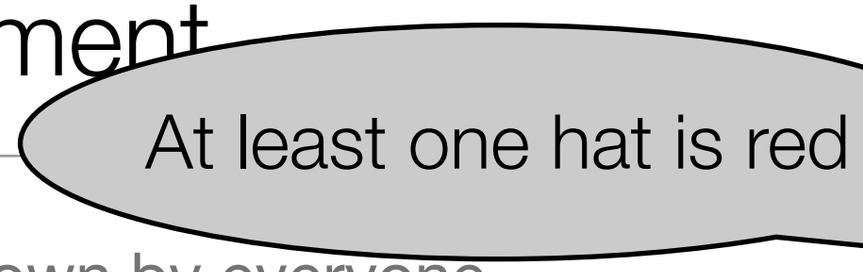
- The teacher announces something that is already known by everyone
- But which of the following statements are true (before the announcement)?
  - Ann knows that Bill knows that at least one hat is red YES
  - Bill knows that Ann knows that at least one hat is red YES
  - Ann knows that Bill knows that Cath knows that at least one hat is red

The role of the teacher's announcement



- The teacher announces something that is already known by everyone
- But which of the following statements are true (before the announcement)?
  - Ann knows that Bill knows that at least one hat is red YES
  - Bill knows that Ann knows that at least one hat is red YES
  - Ann knows that Bill knows that Cath knows that at least one hat is red NO

The role of the teacher's announcement



- The teacher announces something that is already known by everyone
- But which of the following statements are true (before the announcement)?
  - Ann knows that Bill knows that at least one hat is red YES
  - Bill knows that Ann knows that at least one hat is red YES
  - Ann knows that Bill knows that Cath knows that at least one hat is red NO
- But all of this (and more) is communicated by the announcement

Common Knowledge

- Something is common knowledge if everybody knows it, everybody knows that everybody knows it, and so on (ad infinitum)

- First known from the theory of *conventions* (Lewis)

- Also important in game theory (Aumann)

How is common knowledge created?

- *Not* by learning each of (infinitely many) propositions “everybody knows that ..”

- But by *jointly observing some public event*
  - the teacher’s announcement

- Similar to learning a universally quantified first-order formula

Common knowledge in the 3 Chinese students example

- p: “at least one hat is red”

- From the teacher’s announcement, each of the students learn that “it is common knowledge that p” - and this was not the case before the announcement

- From that, Ann can *deduce* that
  - “p”
  - “Bill knows p”
  - “Bill knows that Cath knows p”
  - ...

Reasoning about Knowledge: Formalisation

- Propositional logic: not enough
- First-order (predicate) logic
    - Knows(Ann,p)
    - Used in AI in the 80s
    - Some technical complications: reification, etc.
- **Modal logic**
    - From the 90s, now the most popular approach
    - Well suited for *propositional attitudes*
    - Computational advantages

Modal logic for reasoning about knowledge

Language: extends propositional logic with *knowledge operators*

$$\phi ::= p \mid \neg\phi_1 \mid \phi_1 \wedge \phi_2 \mid K_A\phi$$

Modal logic for reasoning about knowledge

Modal logic for reasoning about knowledge

p : it is raining

Modal logic for reasoning about knowledge

p : it is raining

$K_A p$: Ann knows that it is raining

Modal logic for reasoning about knowledge

p : it is raining

$K_A p$: Ann knows that it is raining

$K_A(p \wedge \neg q)$: Ann knows that it is raining and that it is not Tuesday

Modal logic for reasoning about knowledge

p : it is raining

$K_A p$: Ann knows that it is raining

$K_A(p \wedge \neg q)$: Ann knows that it is raining and that it is not Tuesday

$K_A K_B p$: Ann knows that Bill knows that it is raining

Modal logic for reasoning about knowledge

p : it is raining

$K_A p$: Ann knows that it is raining

$K_A(p \wedge \neg q)$: Ann knows that it is raining and that it is not Tuesday

$K_A K_B p$: Ann knows that Bill knows that it is raining

$K_A(\neg K_B p \wedge \neg K_B \neg p)$: Ann knows that Bill does not know whether it is raining

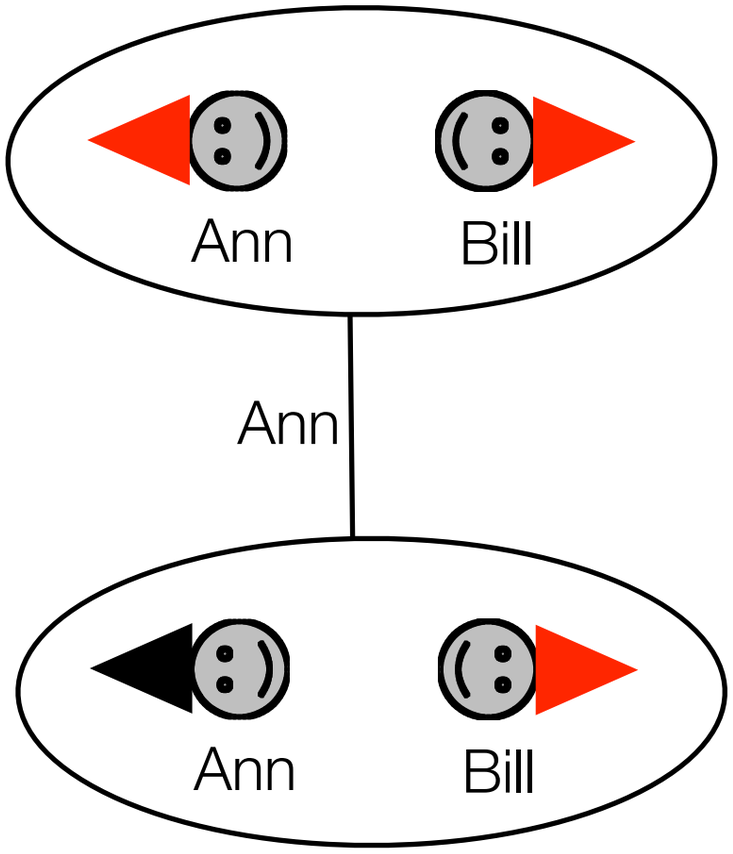
Modal logic for reasoning about knowledge

q : McCord burglered O'Brien's office at Watergate

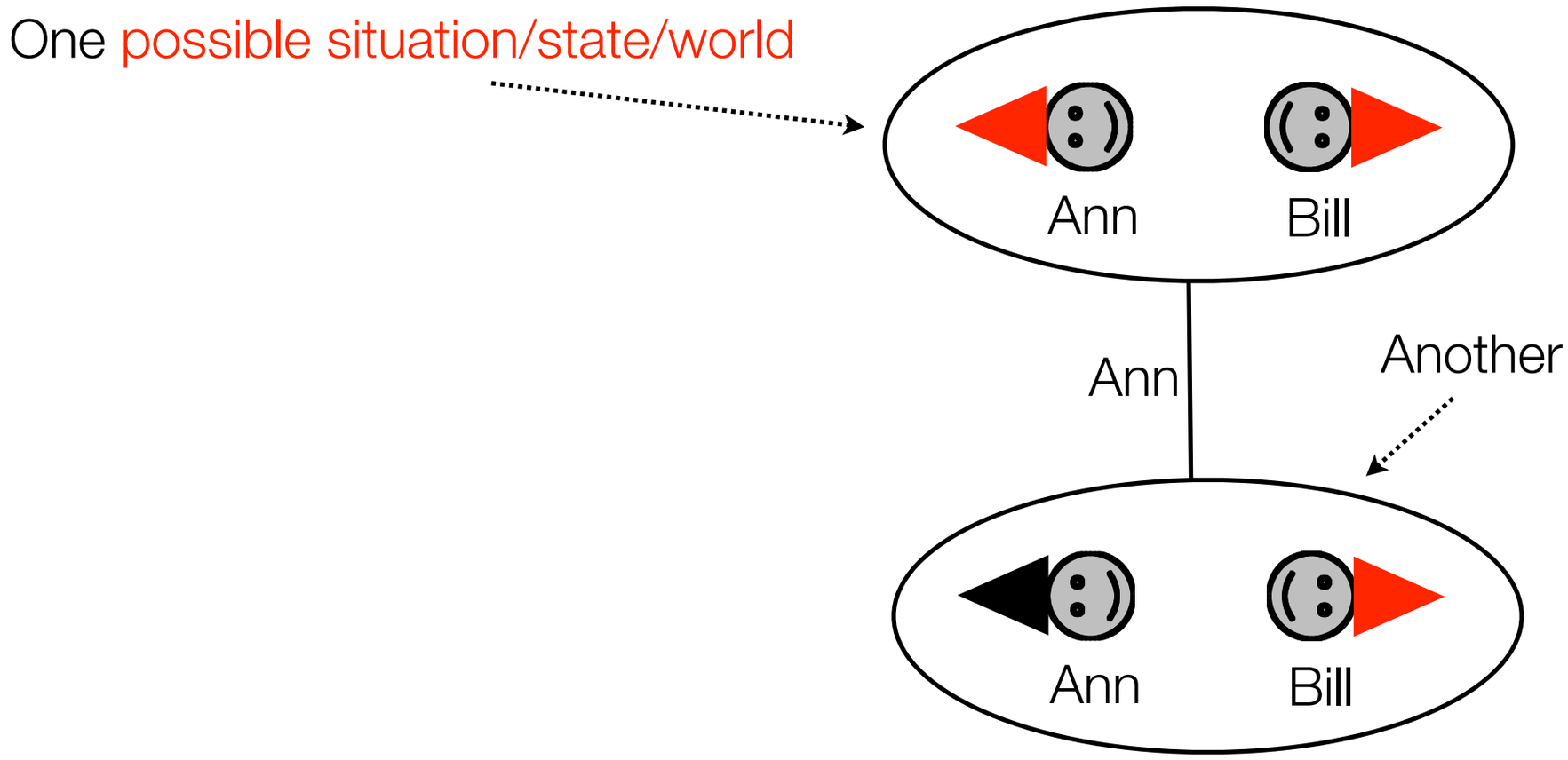
$\neg K_D K_N K_D K_N q \wedge \neg K_D \neg K_N K_D K_N q$

: Dean doesn't know whether Nixon knows that Dean knows that Nixon knows that McCord burgled O'Brien's office at Watergate

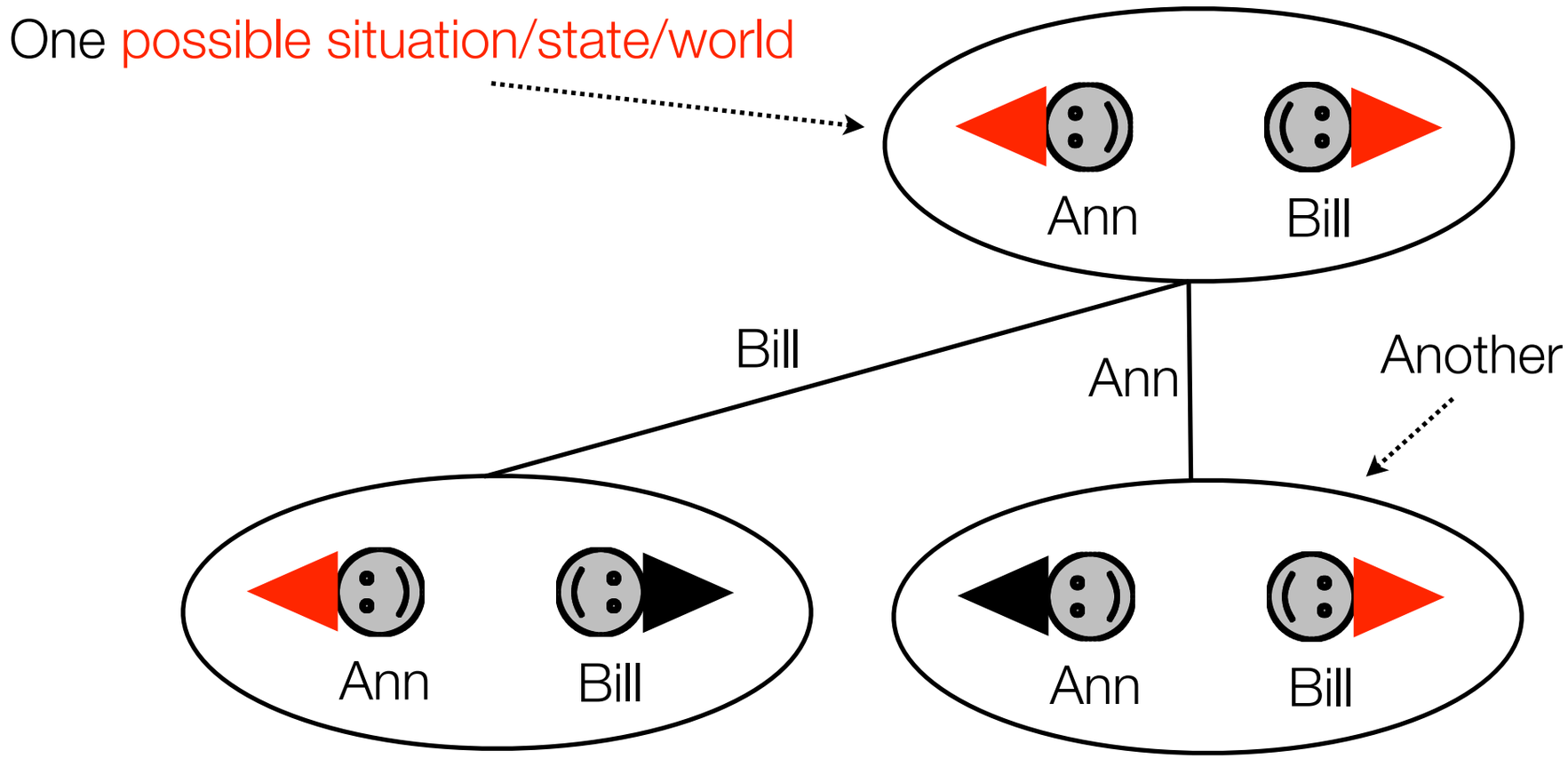
Modal logic: models



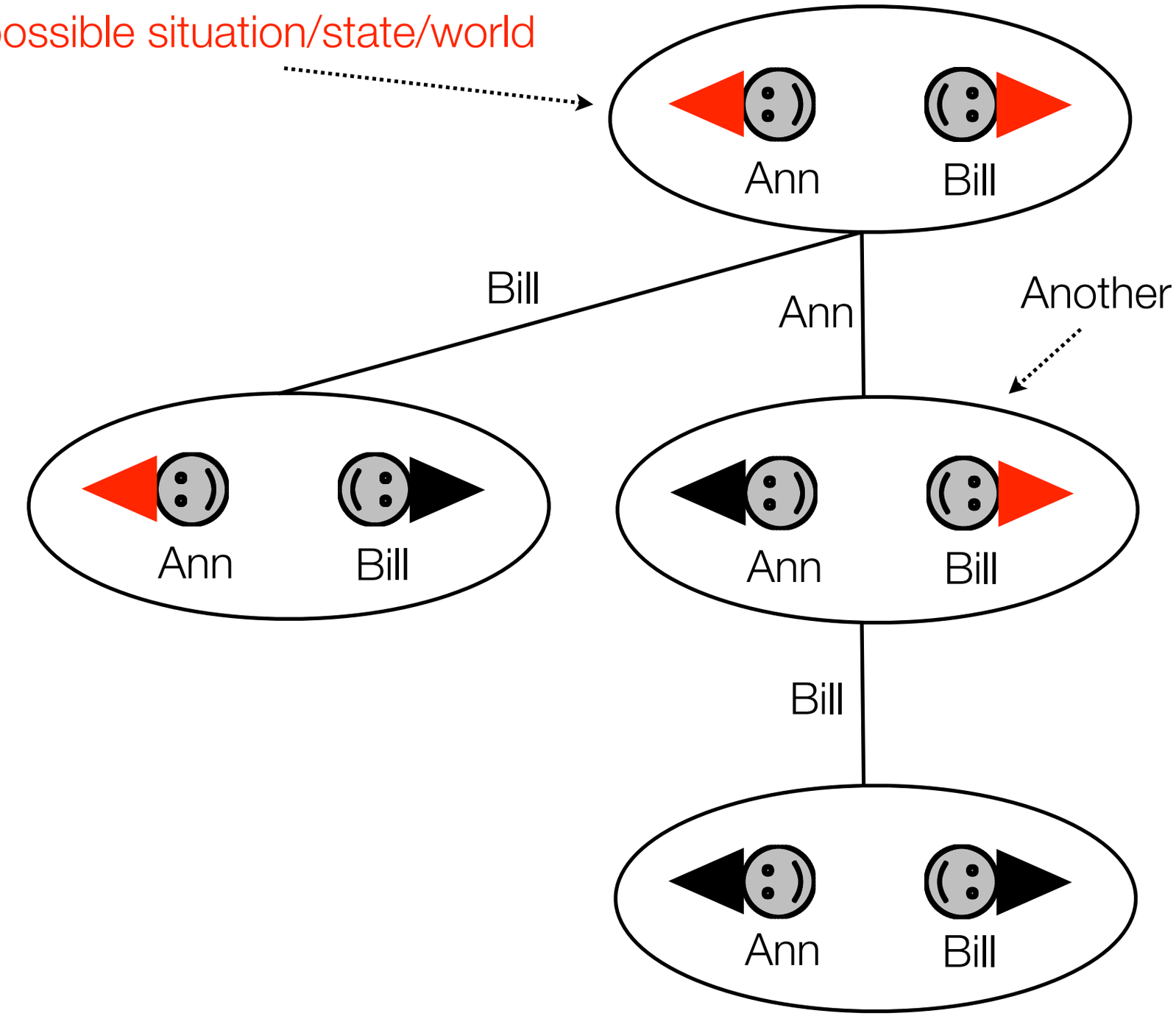
Modal logic: models



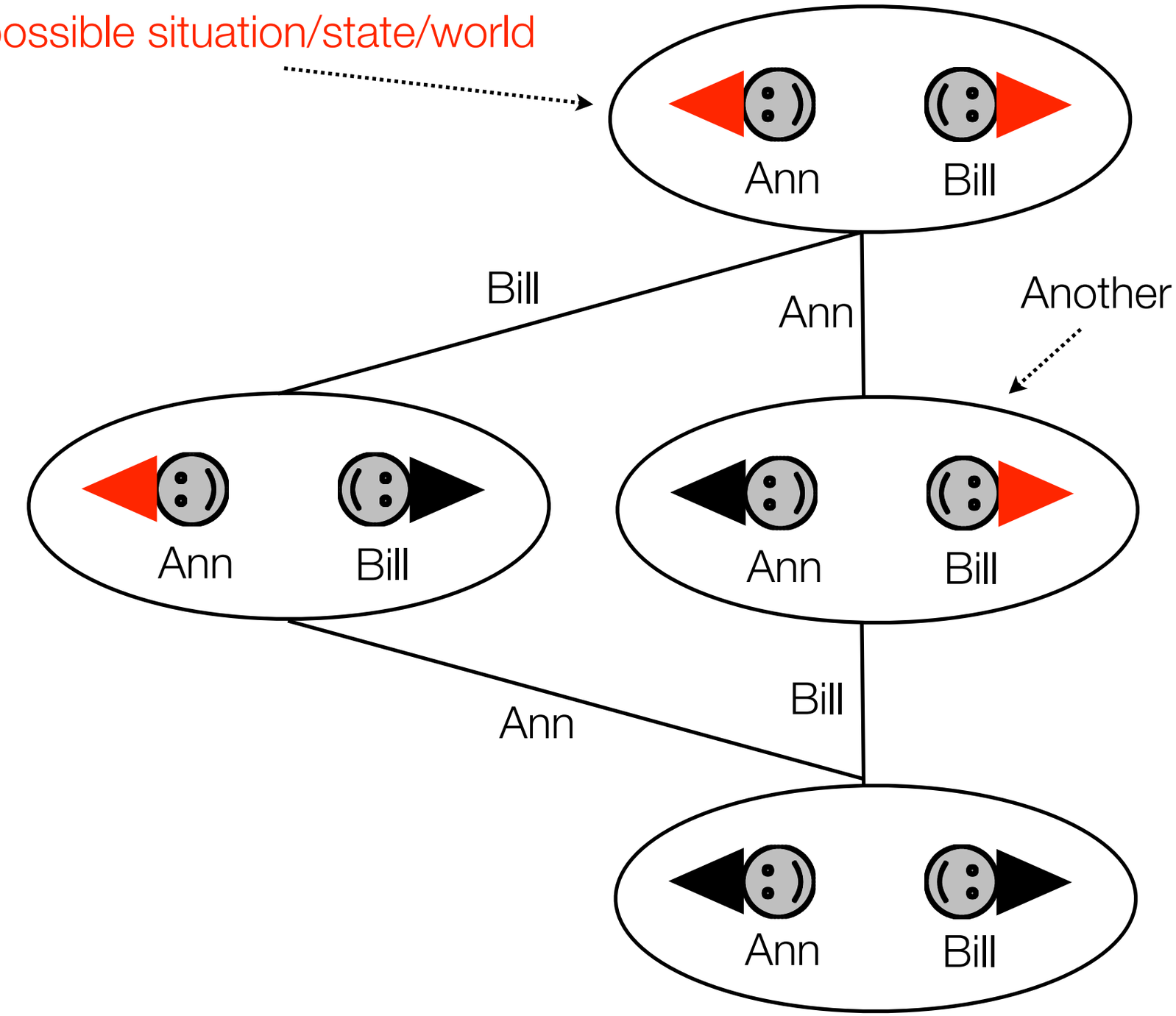
Modal logic: models



Modal logic: models



Modal logic: models

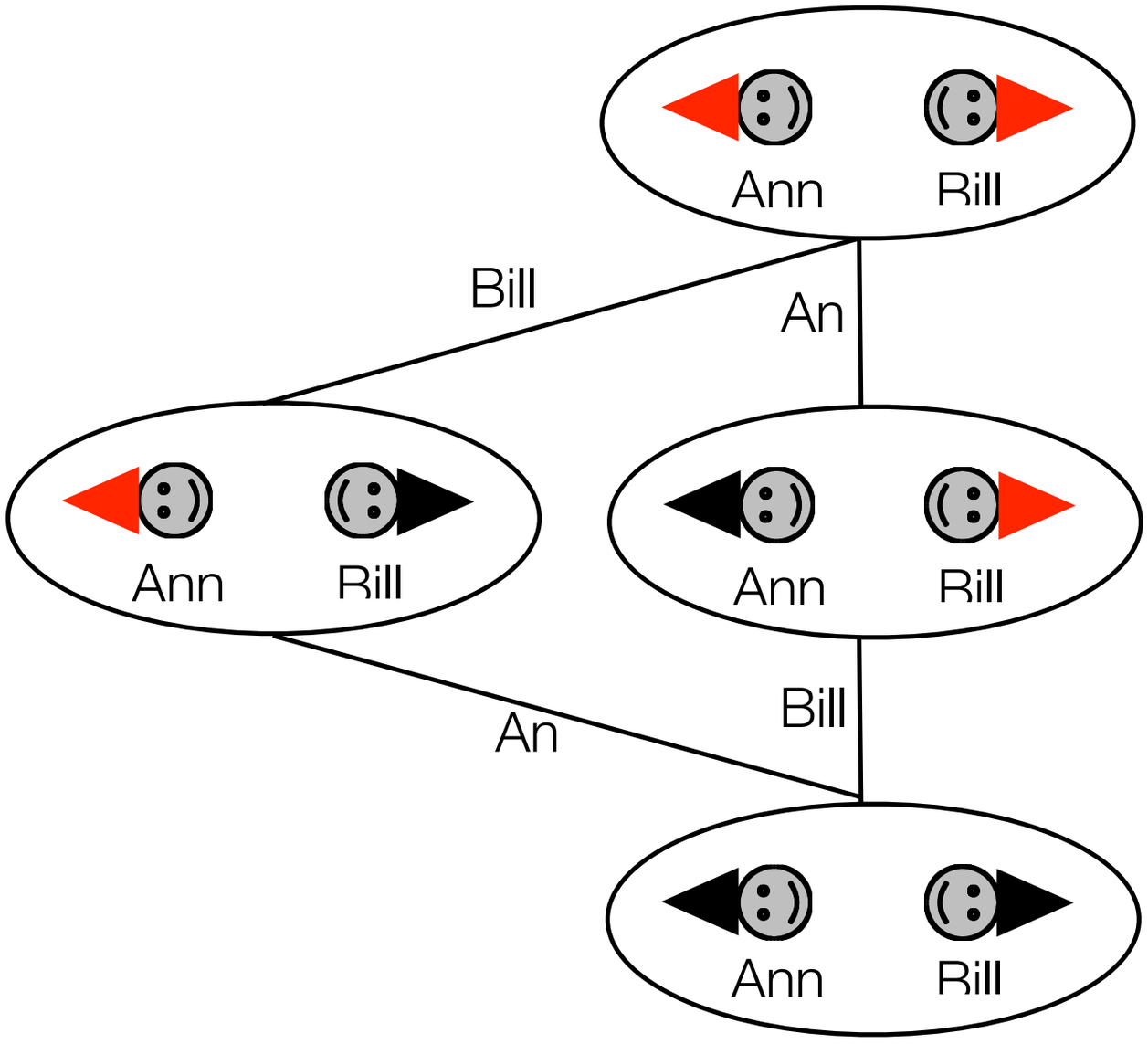


Truth

- Key idea: an agent A knows something in a state if it is true in all states A considers possible

Truth

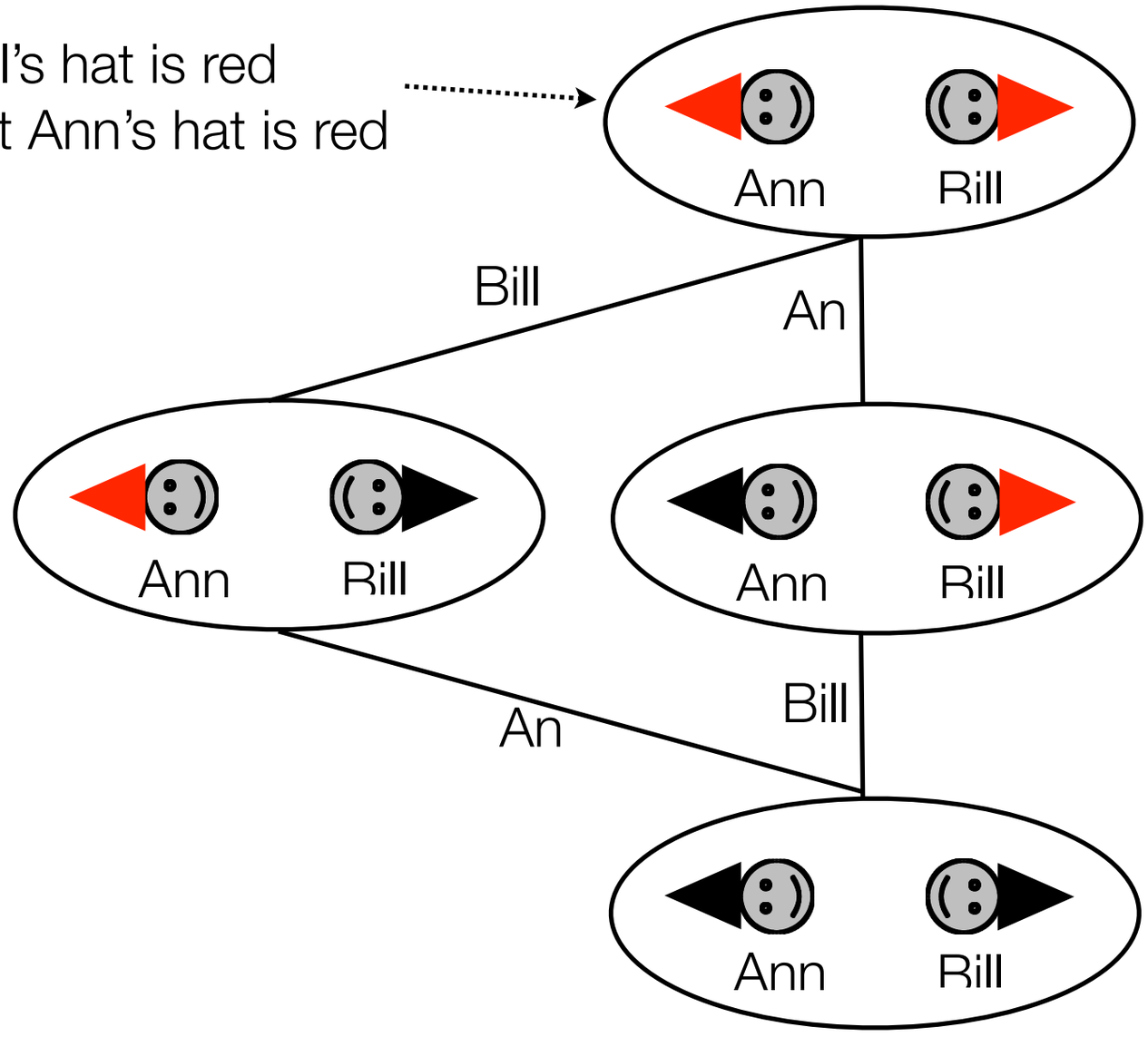
- Key idea: an agent A knows something in a state if it is true in all states A considers possible



Truth

- Key idea: an agent A knows something in a state if it is true in all states A considers possible

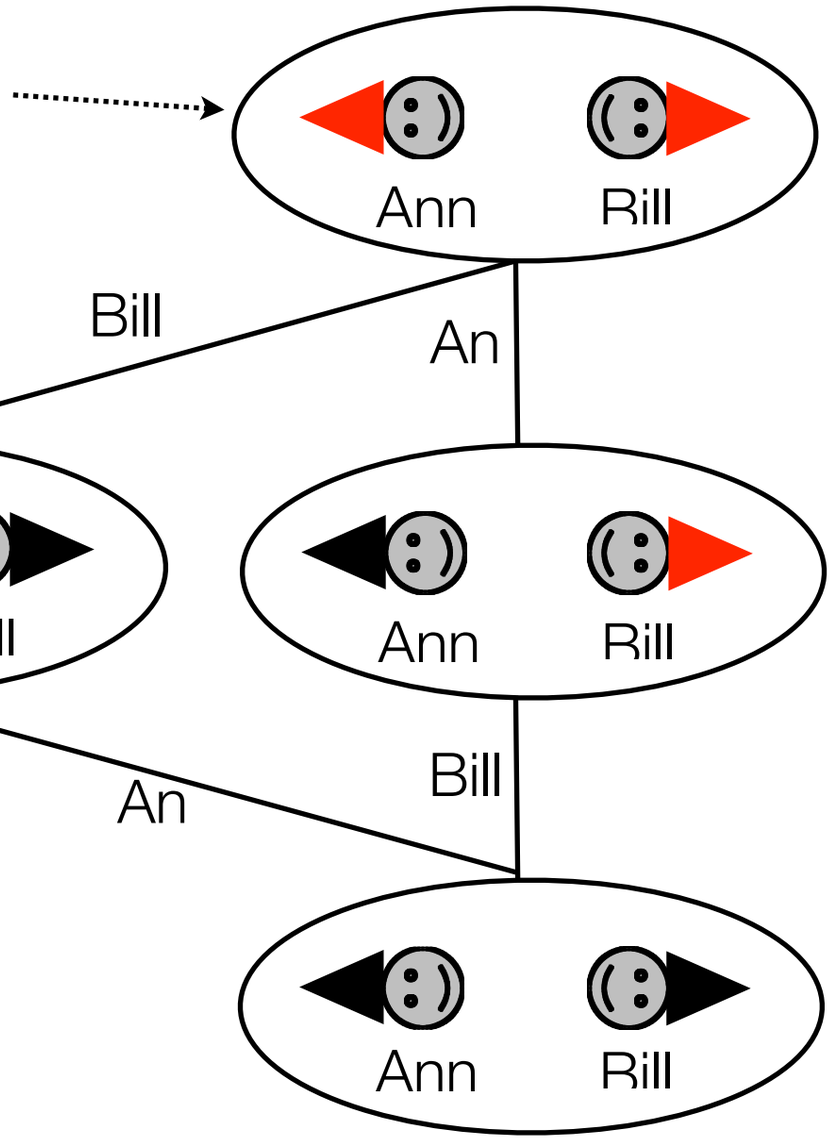
Ann knows that Bill's hat is red
Ann does not know that Ann's hat is red



Truth

- Key idea: an agent A knows something in a state if it is true in all states A considers possible

Ann knows that Bill's hat is red
Ann does not know that Ann's hat is red

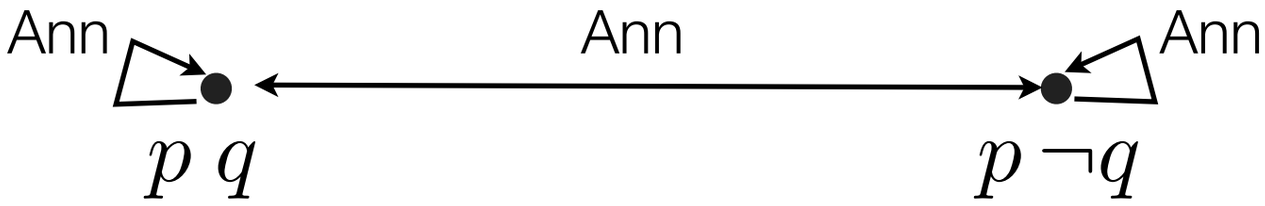


Ann knows that Bill's hat is black
Ann does not know that Ann's hat is red

Semantics: example

p : it is raining

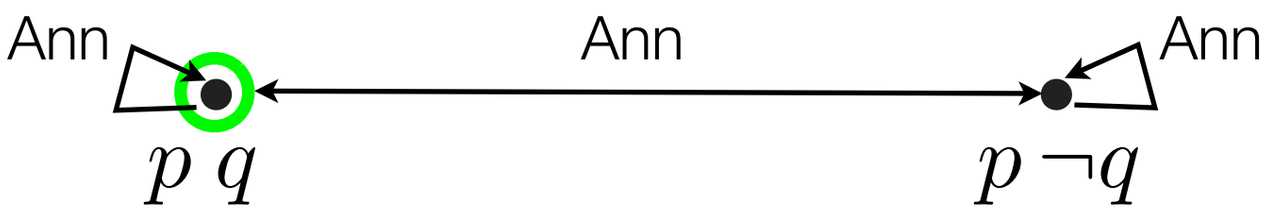
q : it is Wednesday



Semantics: example

p : it is raining

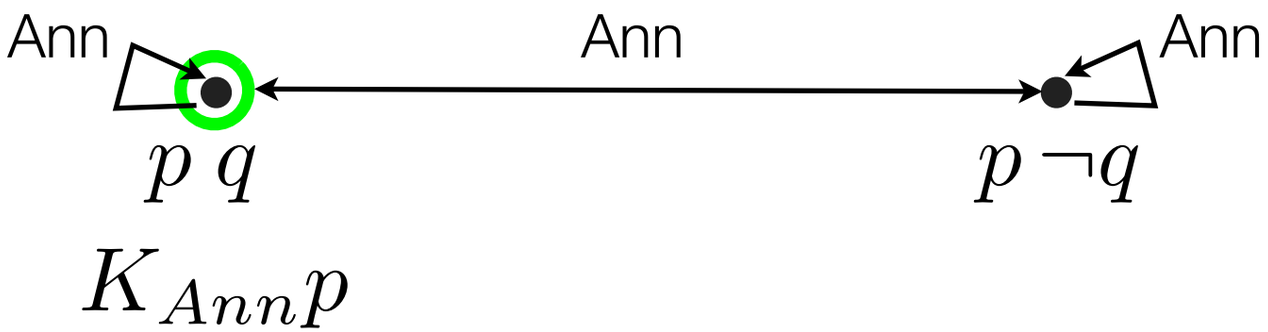
q : it is Wednesday



Semantics: example

p : it is raining

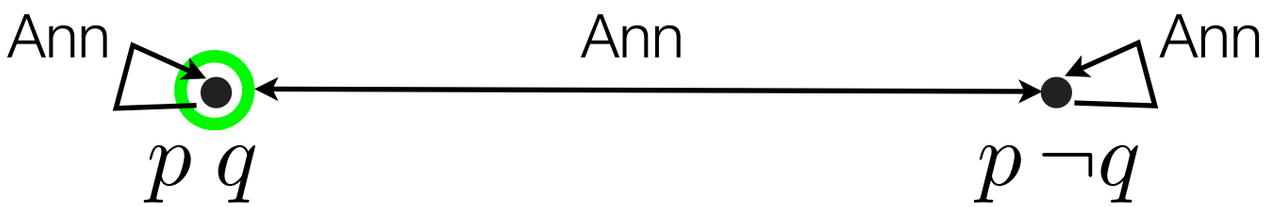
q : it is Wednesday



Semantics: example

p : it is raining

q : it is Wednesday



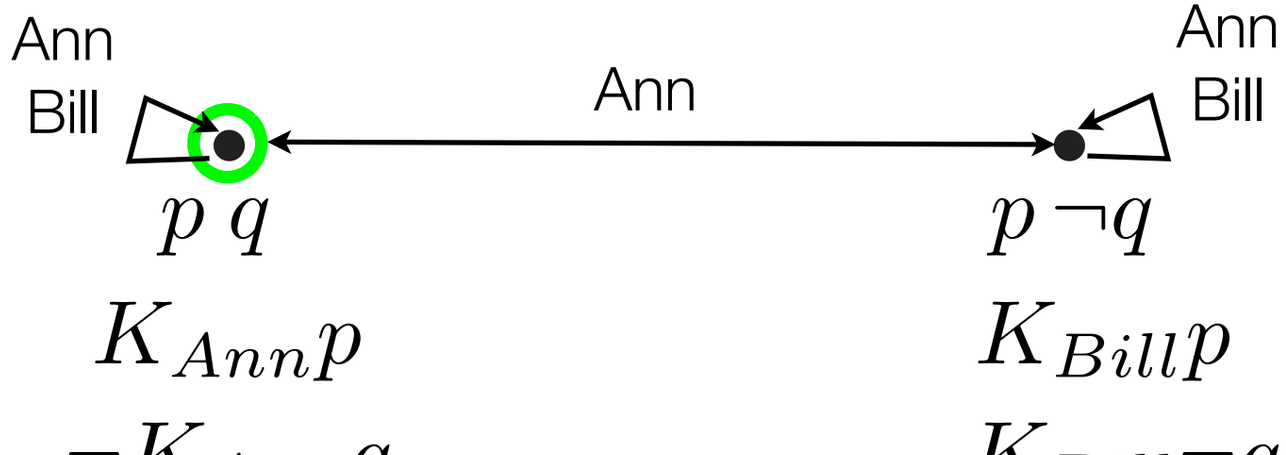
$K_{Ann}p$

$\neg K_{Ann}q$

Semantics: example

p : it is raining

q : it is Wednesday



$p\ q$

$K_{Ann}p$

$\neg K_{Ann}q$

$K_{Bill}p$

$K_{Ann}K_{Bill}p$

$K_{Bill}q$

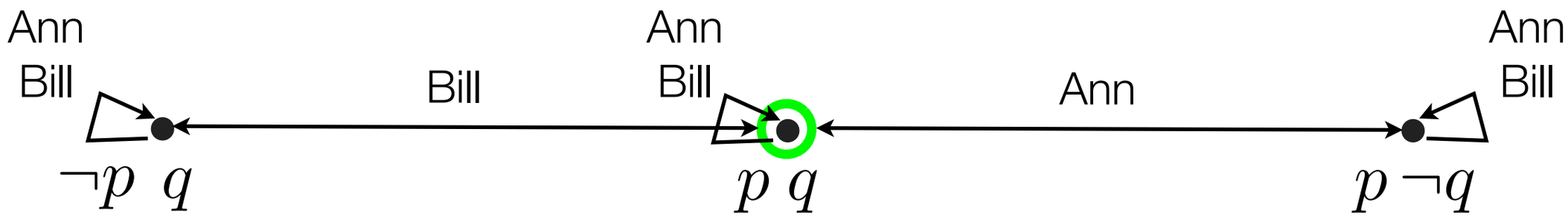
$\neg K_{Ann}K_{Bill}q$

$p\ \neg q$

$K_{Bill}\neg q$

$\neg K_{Bill}q$

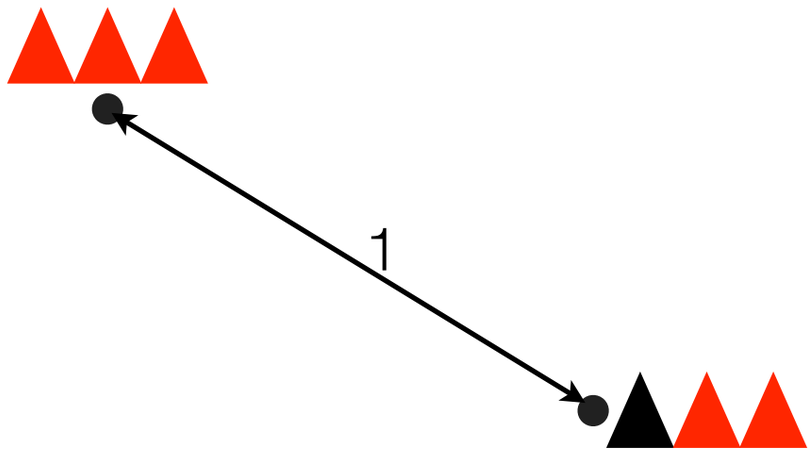
Semantics: example



3 Chinese students

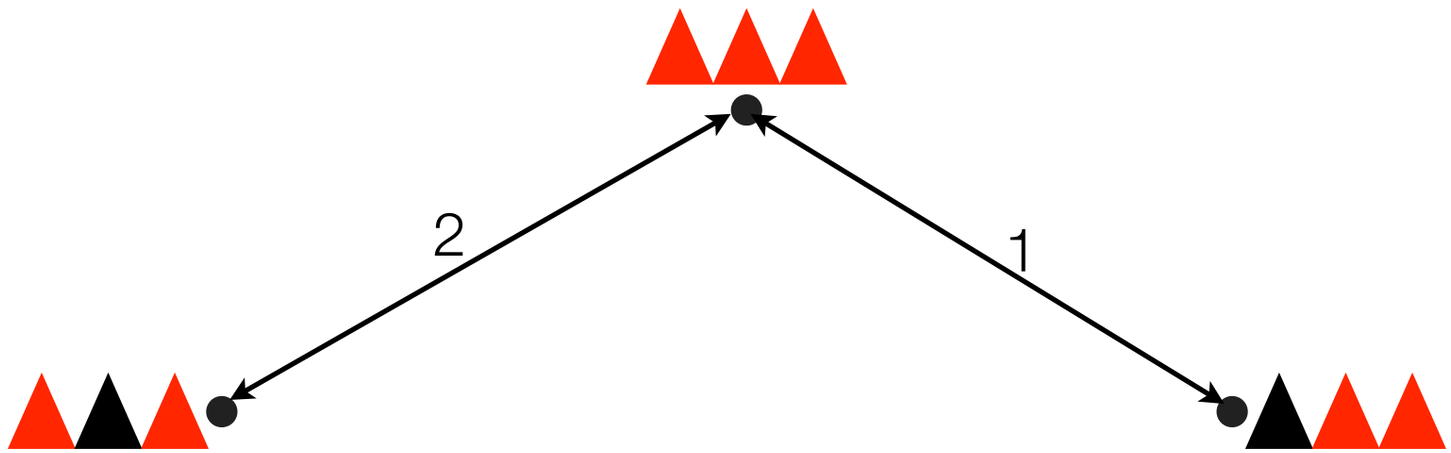
1=Ann, 2=Bill, 3=Cath

3 Chinese students



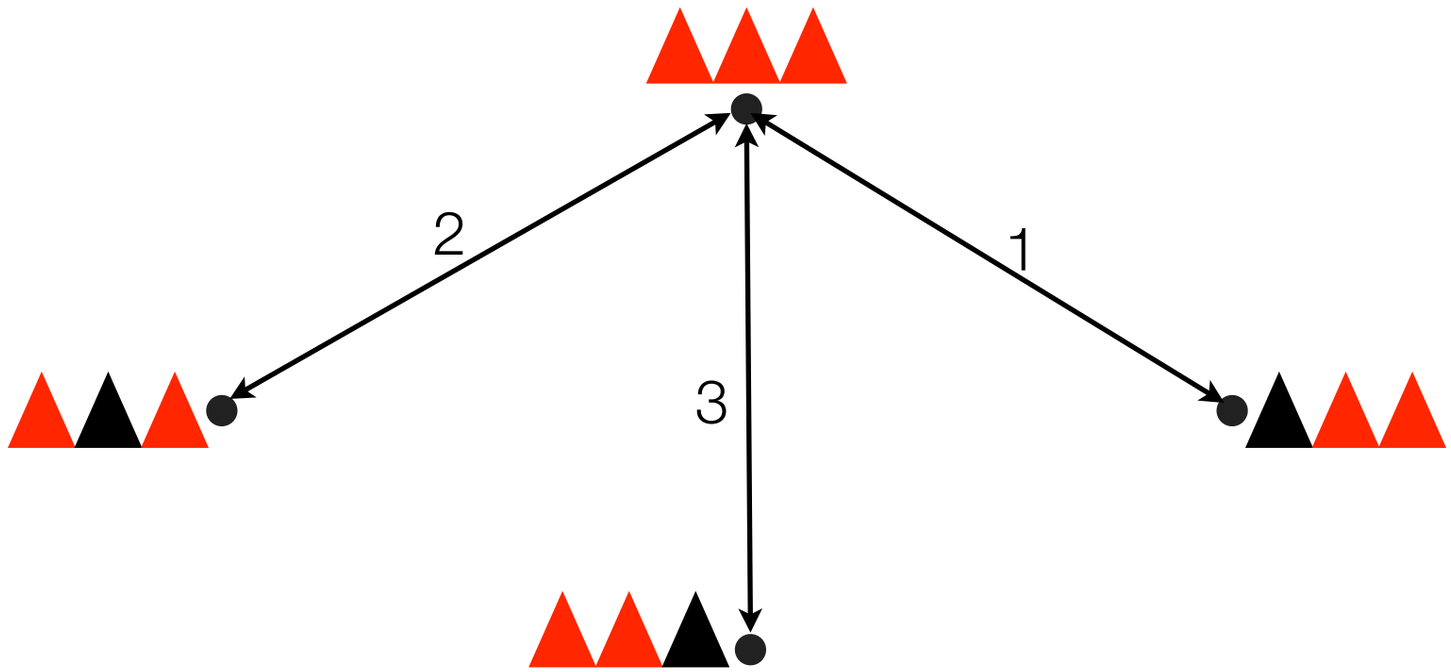
1=Ann, 2=Bill, 3=Cath

3 Chinese students



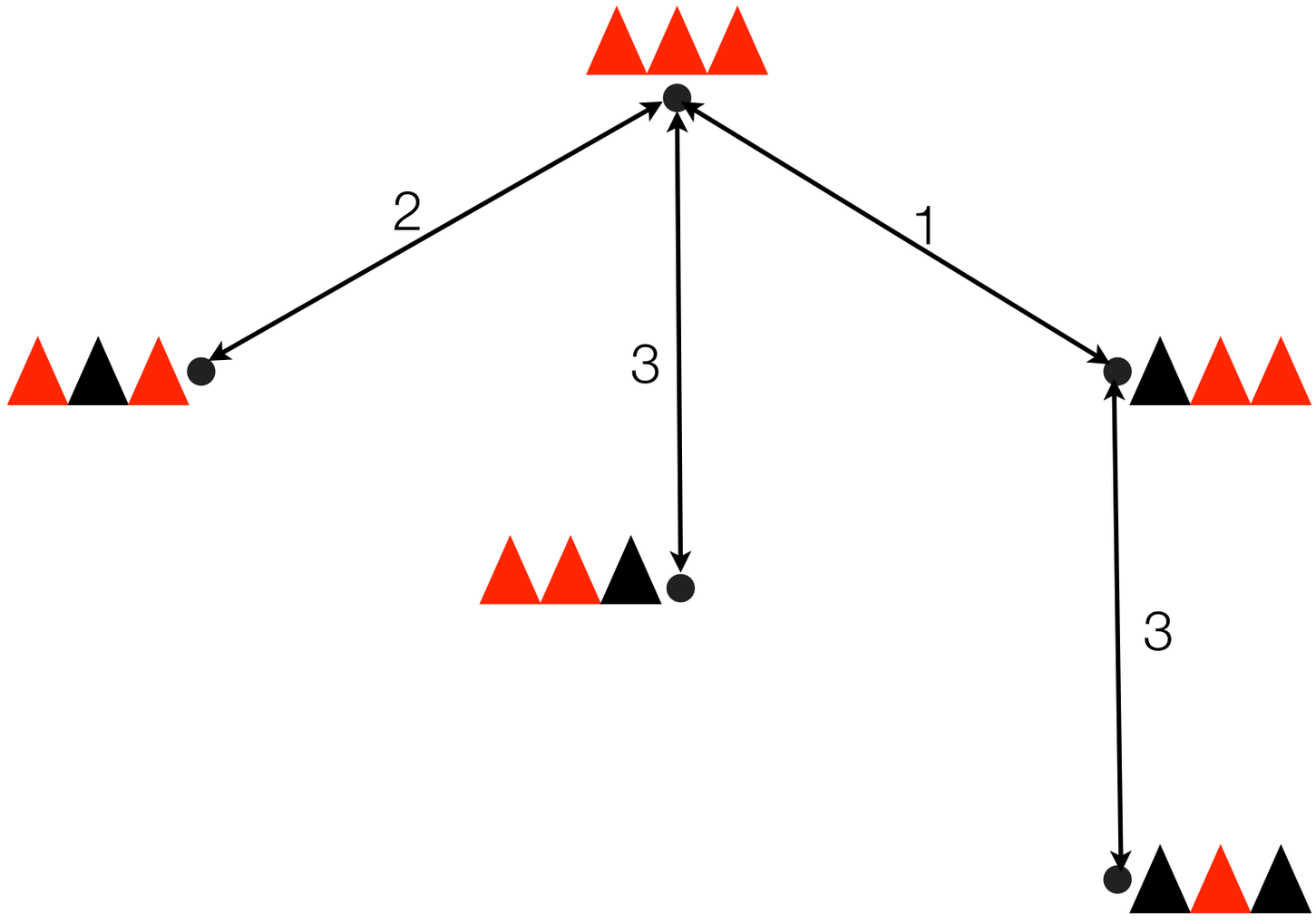
1=Ann, 2=Bill, 3=Cath

3 Chinese students



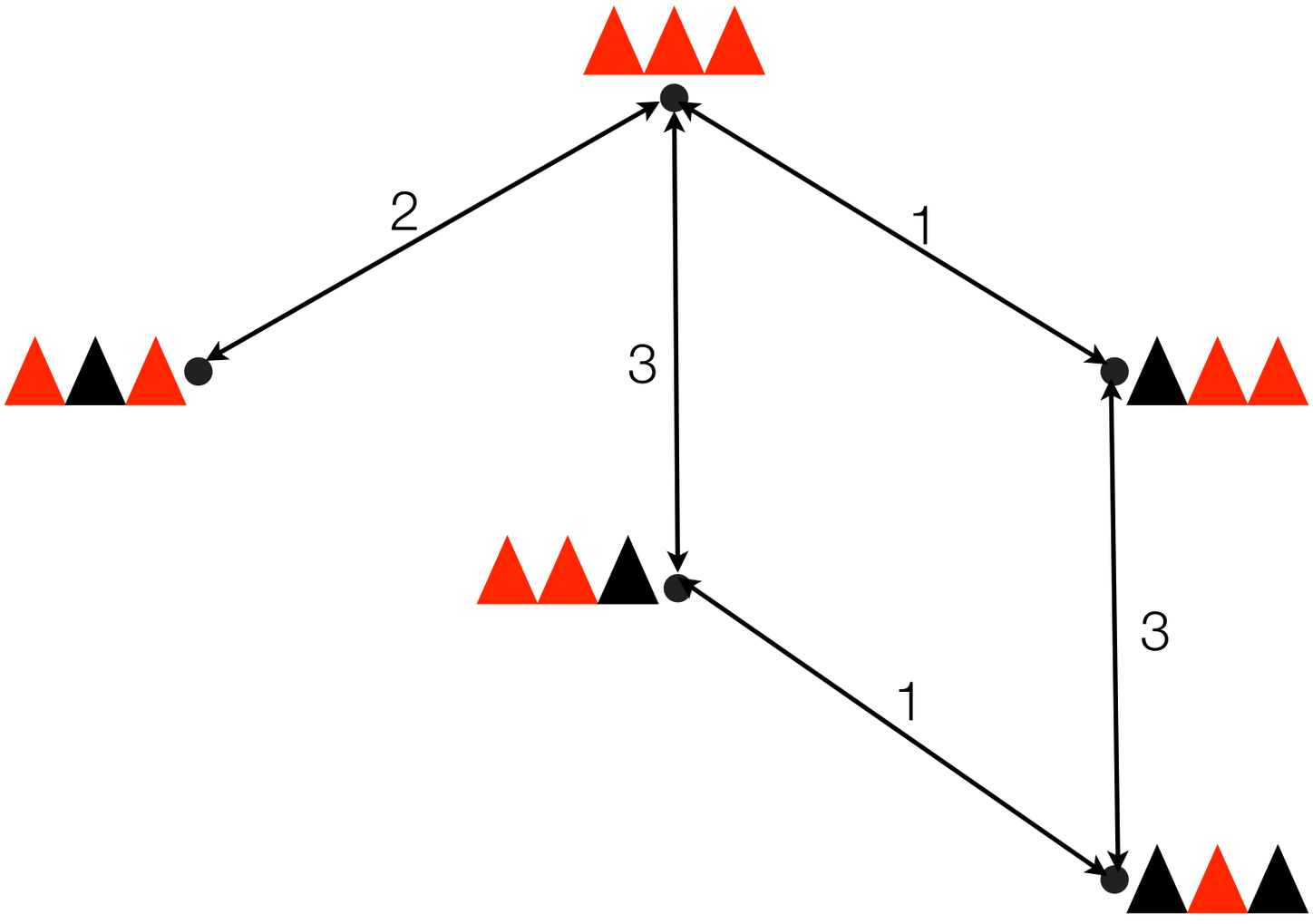
1=Ann, 2=Bill, 3=Cath

3 Chinese students



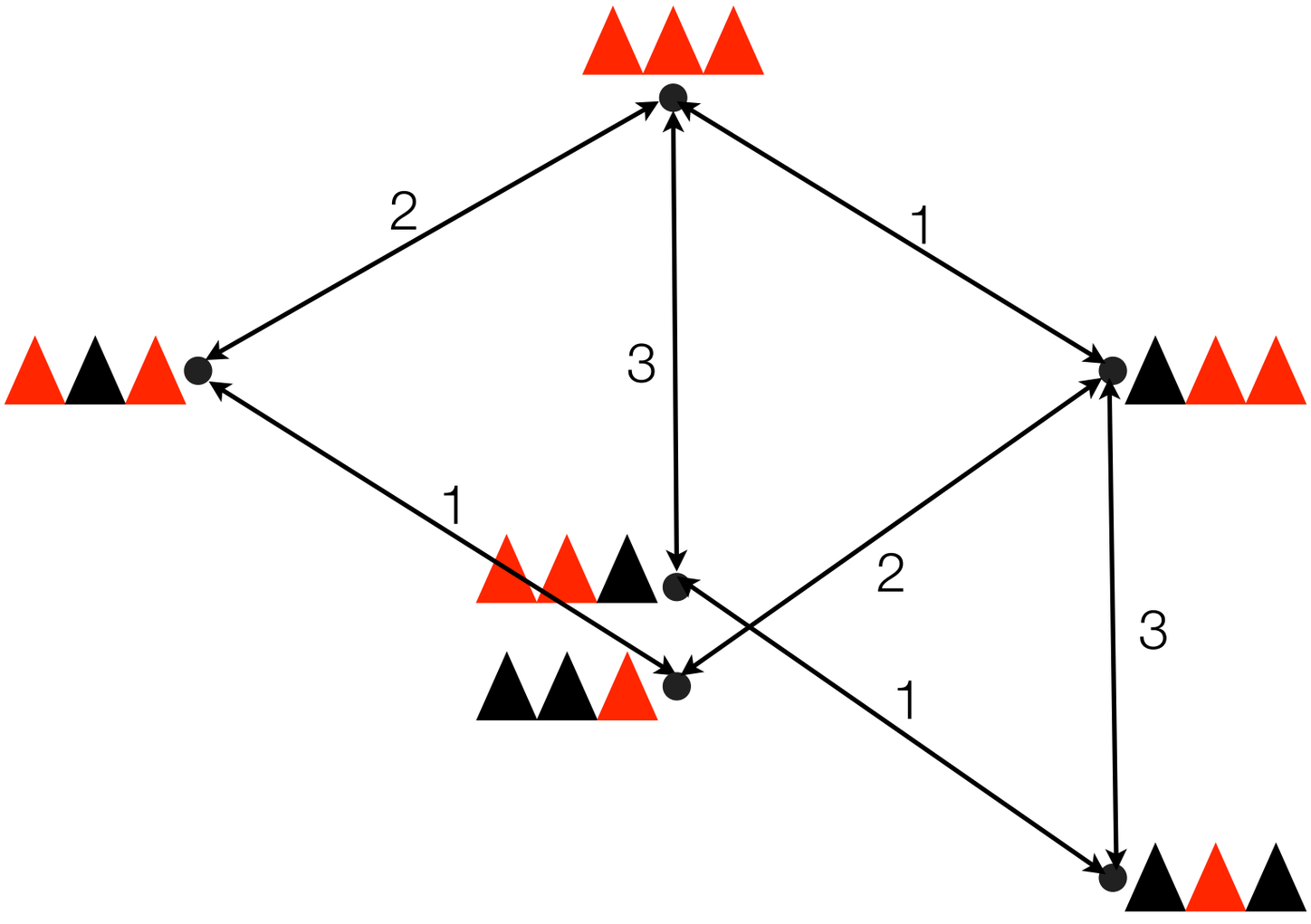
1=Ann, 2=Bill, 3=Cath

3 Chinese students



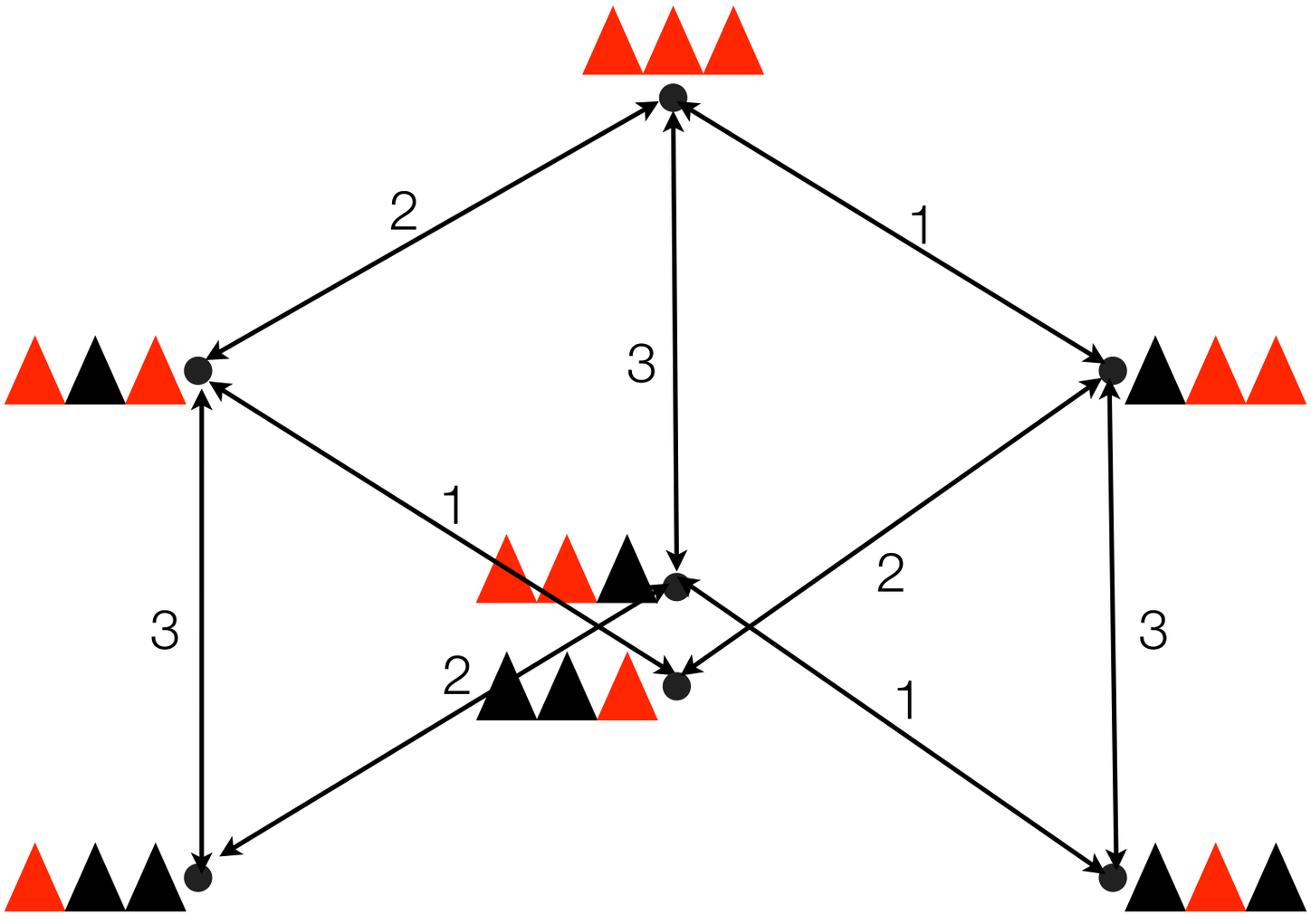
1=Ann, 2=Bill, 3=Cath

3 Chinese students



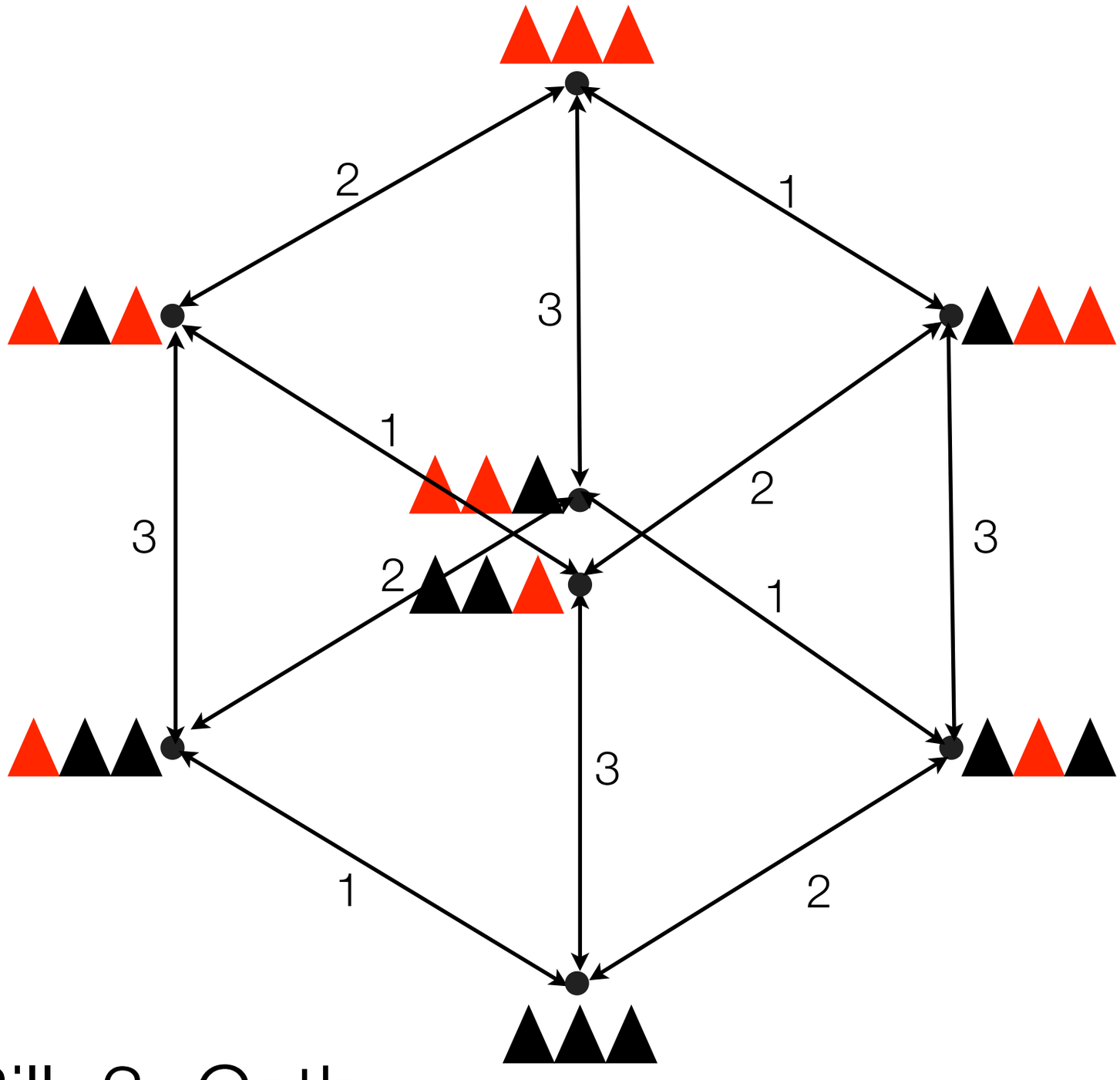
1=Ann, 2=Bill, 3=Cath

3 Chinese students



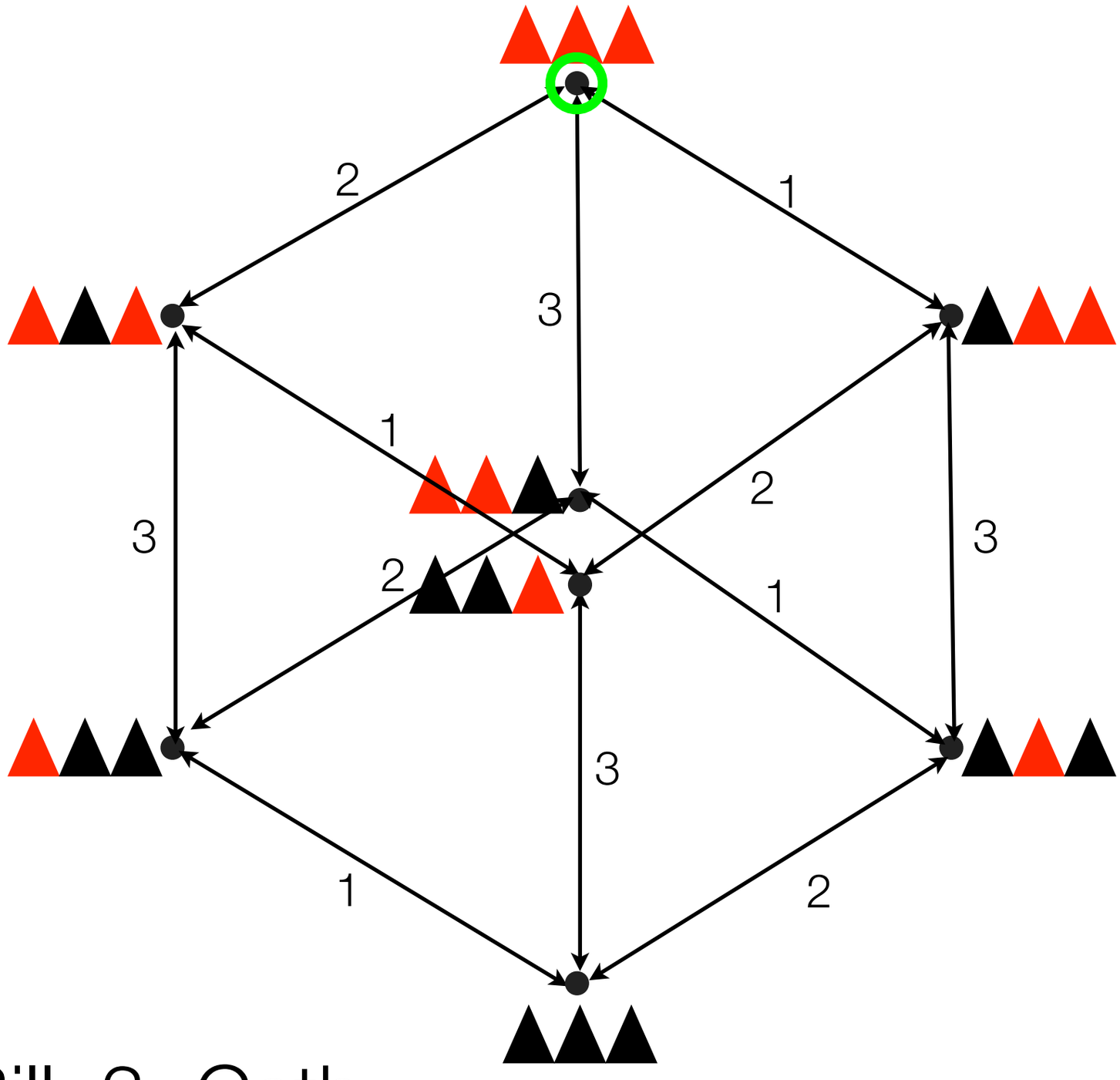
1=Ann, 2=Bill, 3=Cath

3 Chinese students



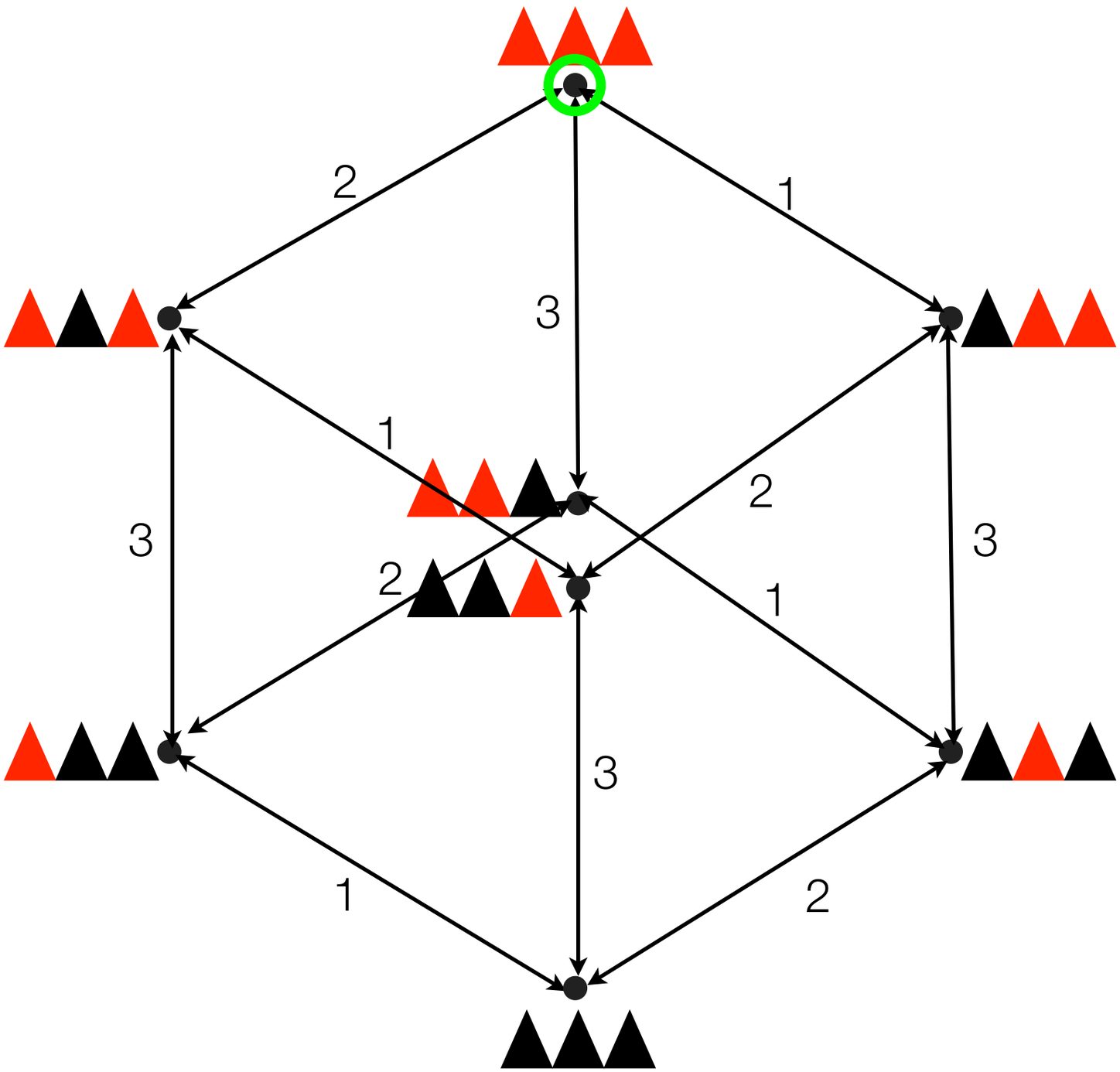
1=Ann, 2=Bill, 3=Cath

3 Chinese students

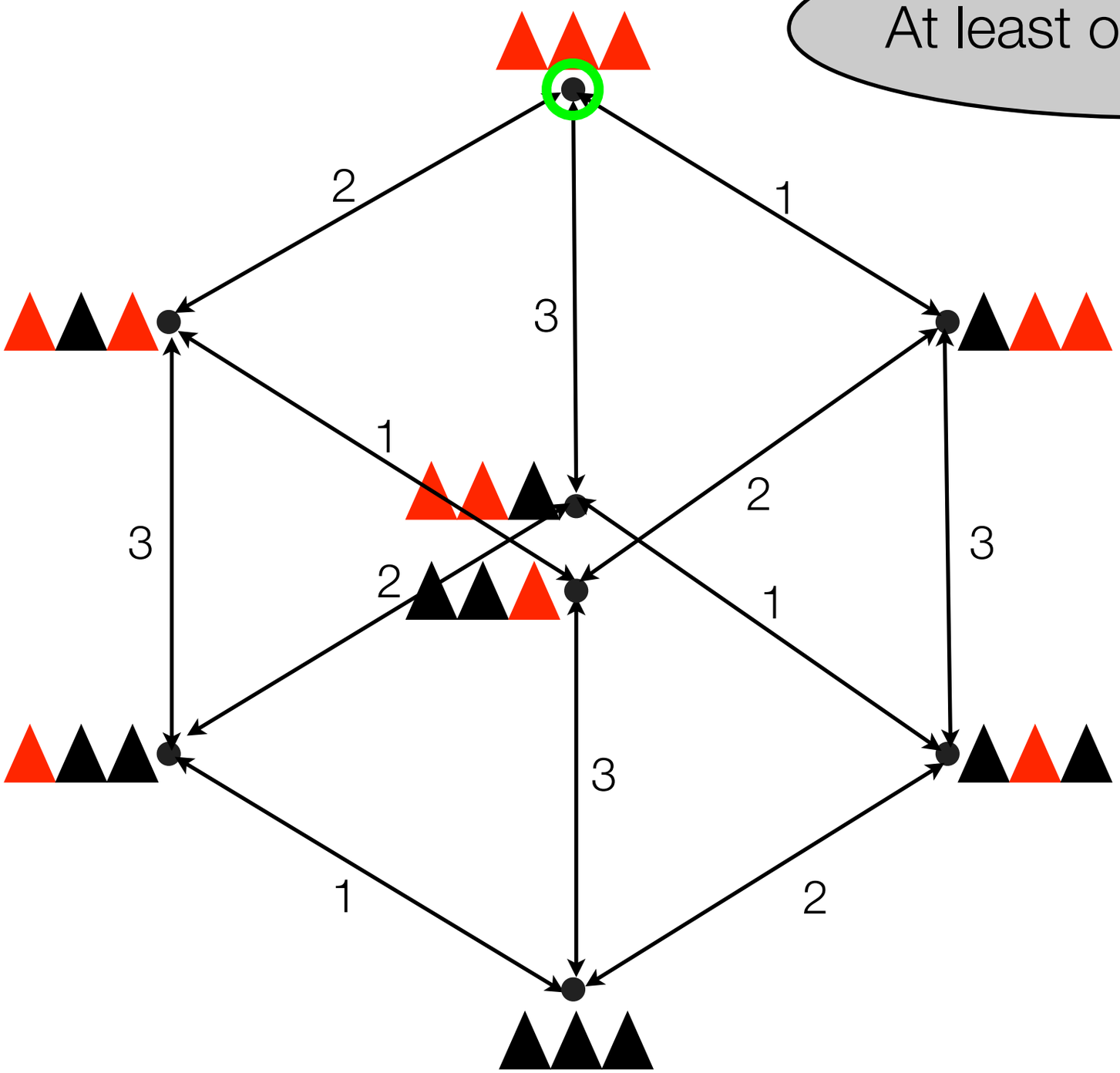


1=Ann, 2=Bill, 3=Cath

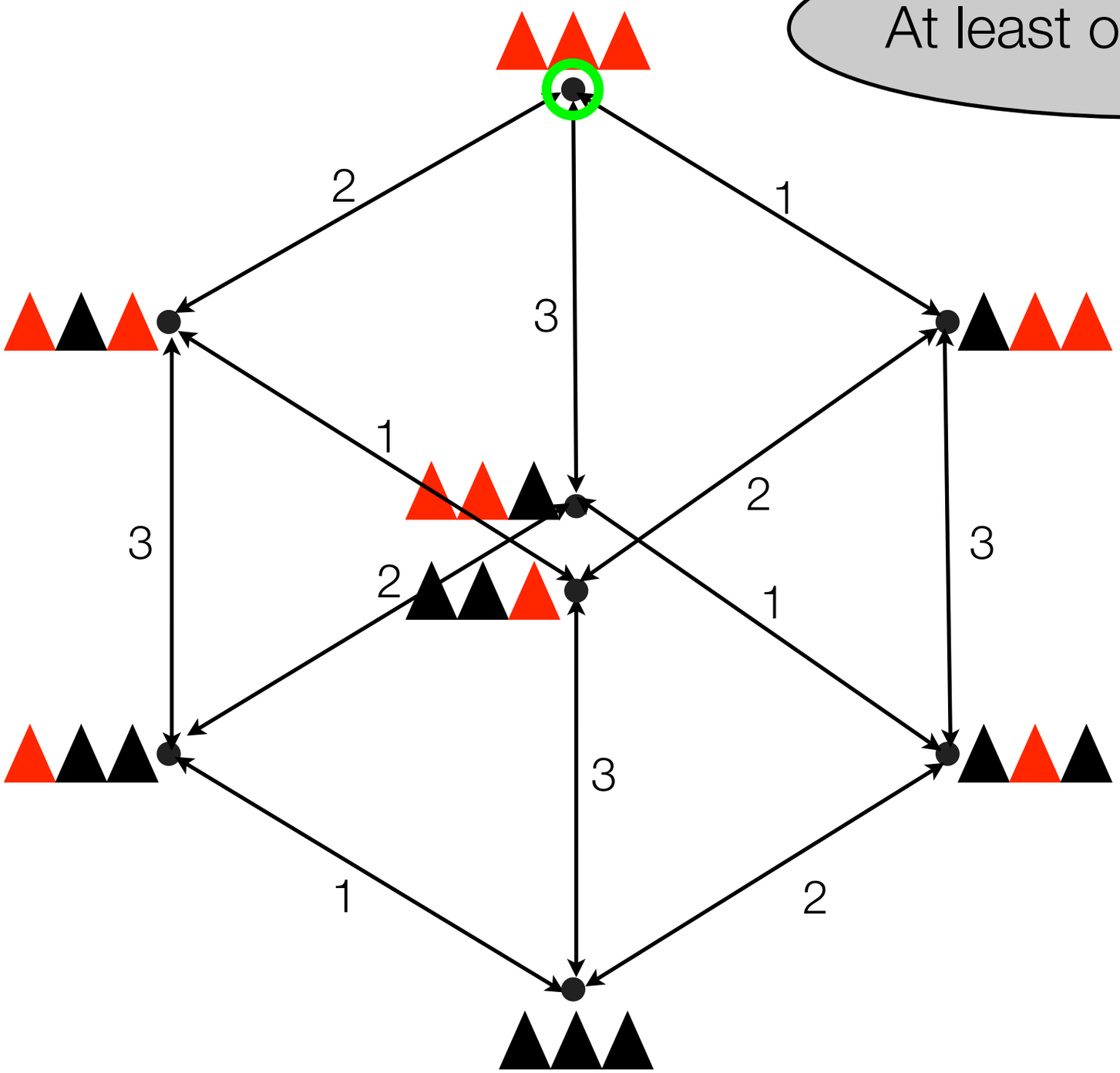
3 Chinese students



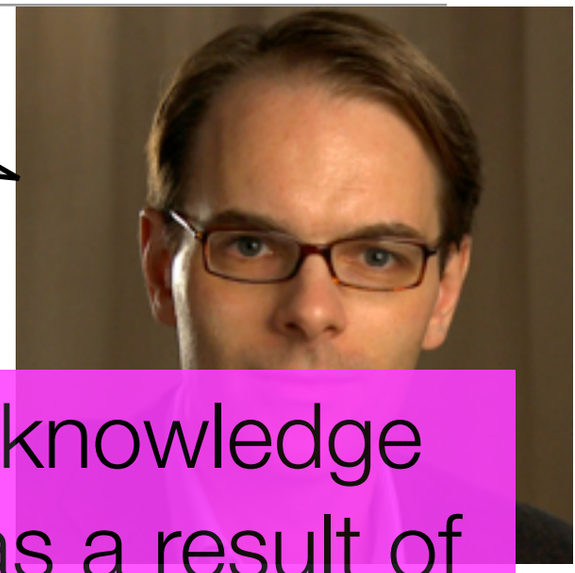
3 Chinese students



3 Chinese students



At least one hat is red



The agents' knowledge is changed as a result of the announcement

How do we change the model to reflect this?

We know: the announcement leads to common knowledge

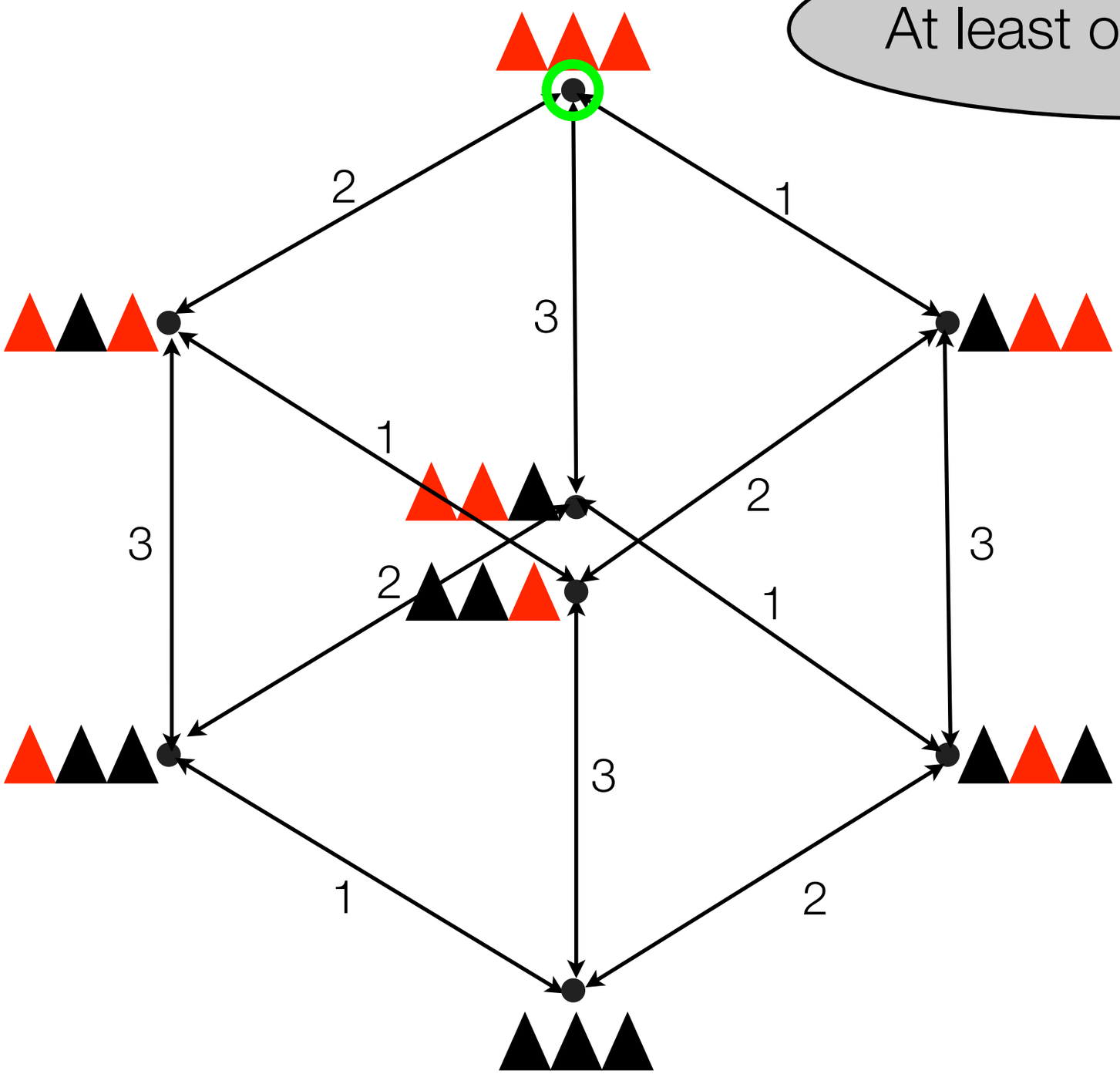
Updating a model after a public announcement

- **p**: the statement that is announced

- **Key idea**: after **p** is announced, no-one considers states where **p** is not true possible any more - those states are *eliminated*

- Delete all arrows pointing to states where **p** is not true

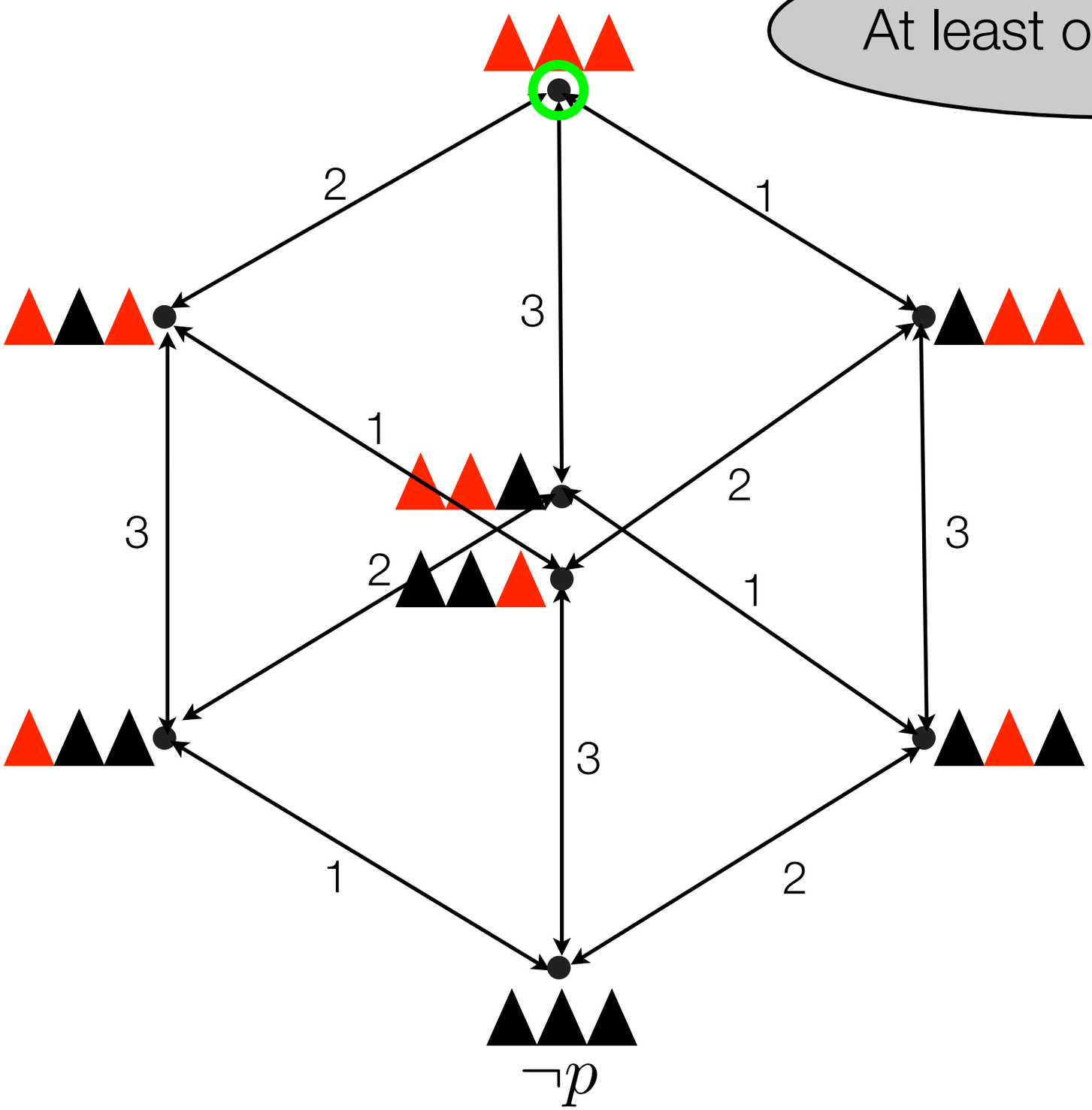
3 Chinese students



At least one hat is red

p

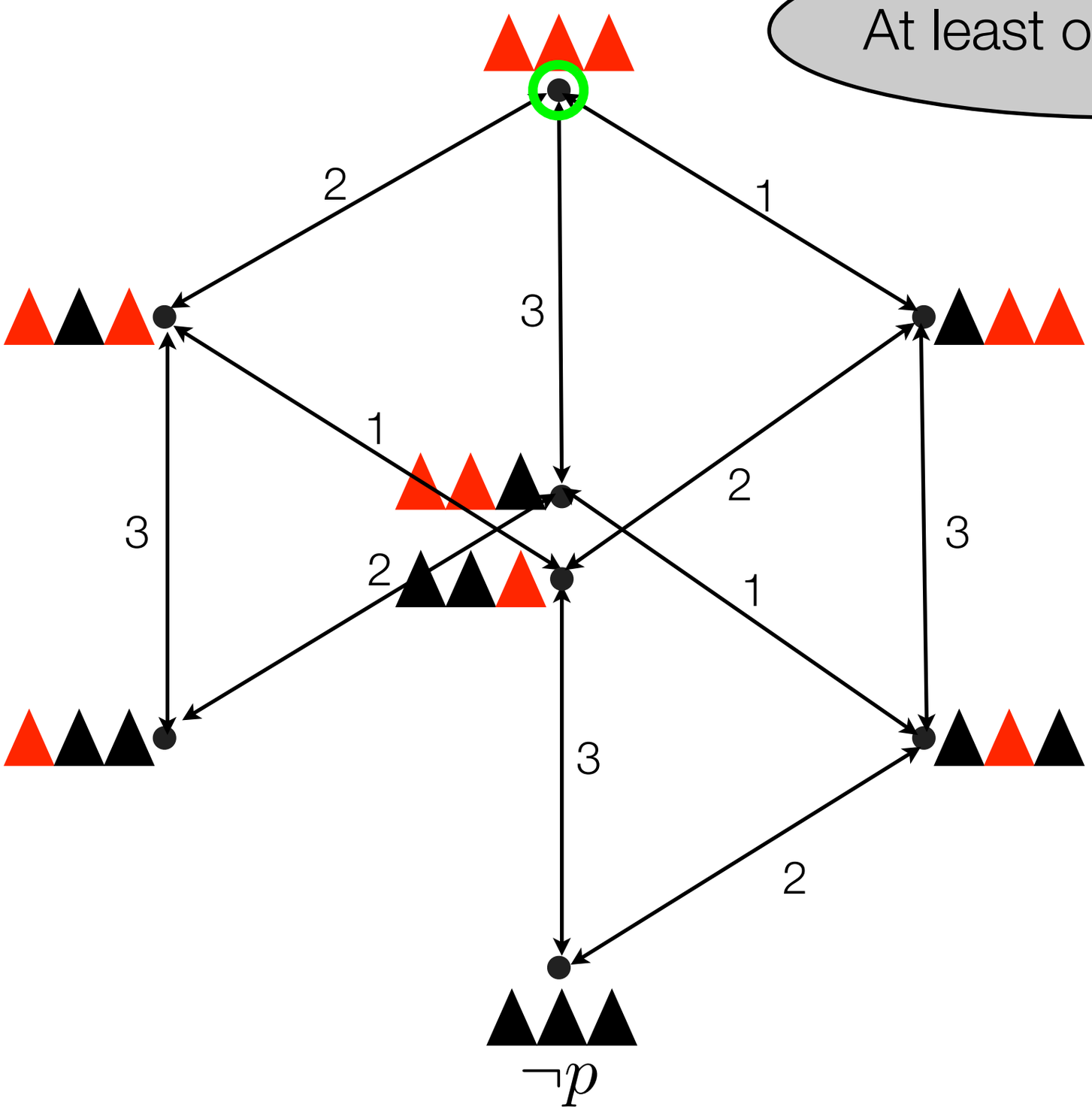
3 Chinese students



At least one hat is red

p

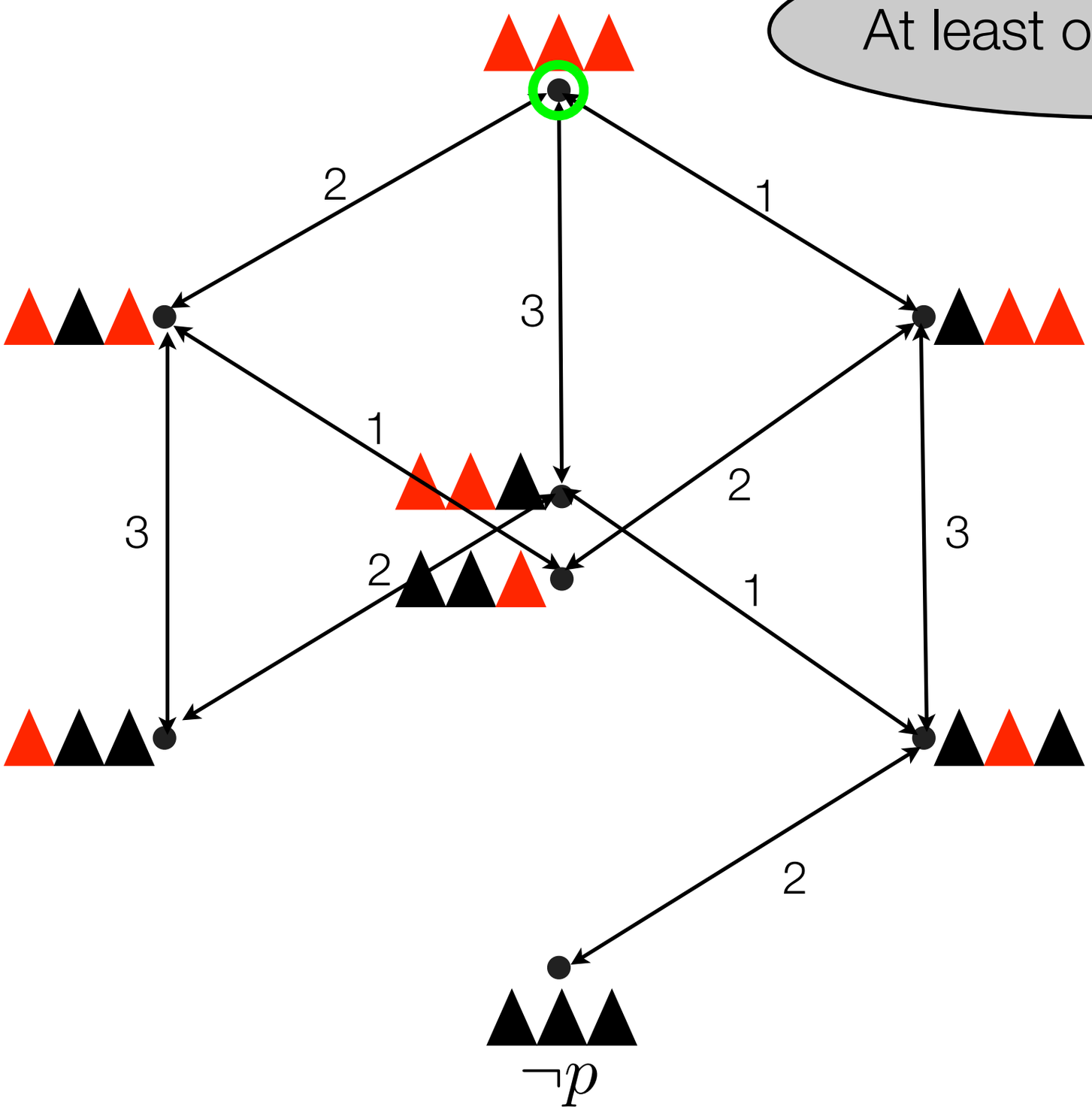
3 Chinese students



At least one hat is red

p

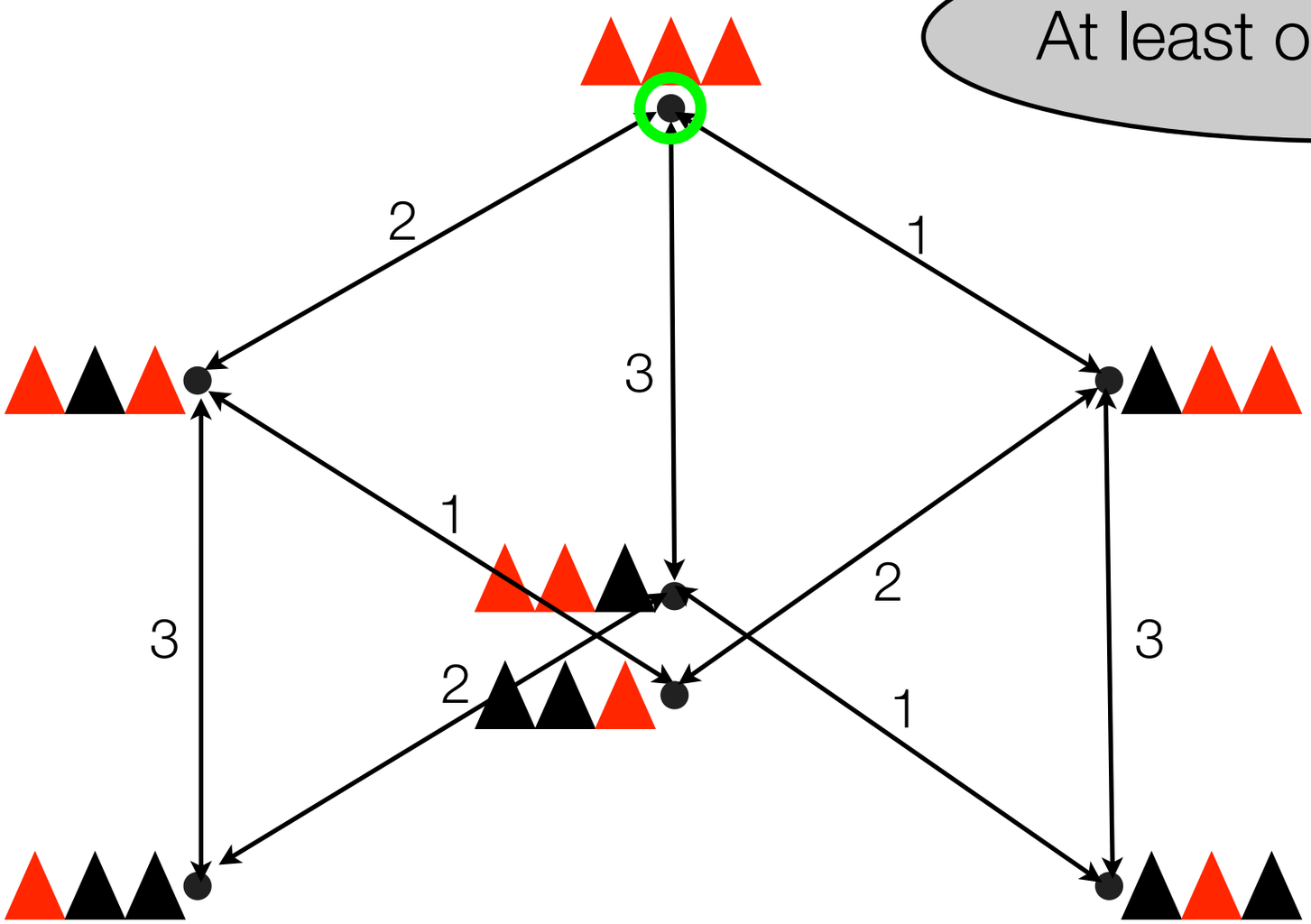
3 Chinese students



At least one hat is red

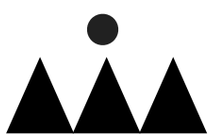
p

3 Chinese students



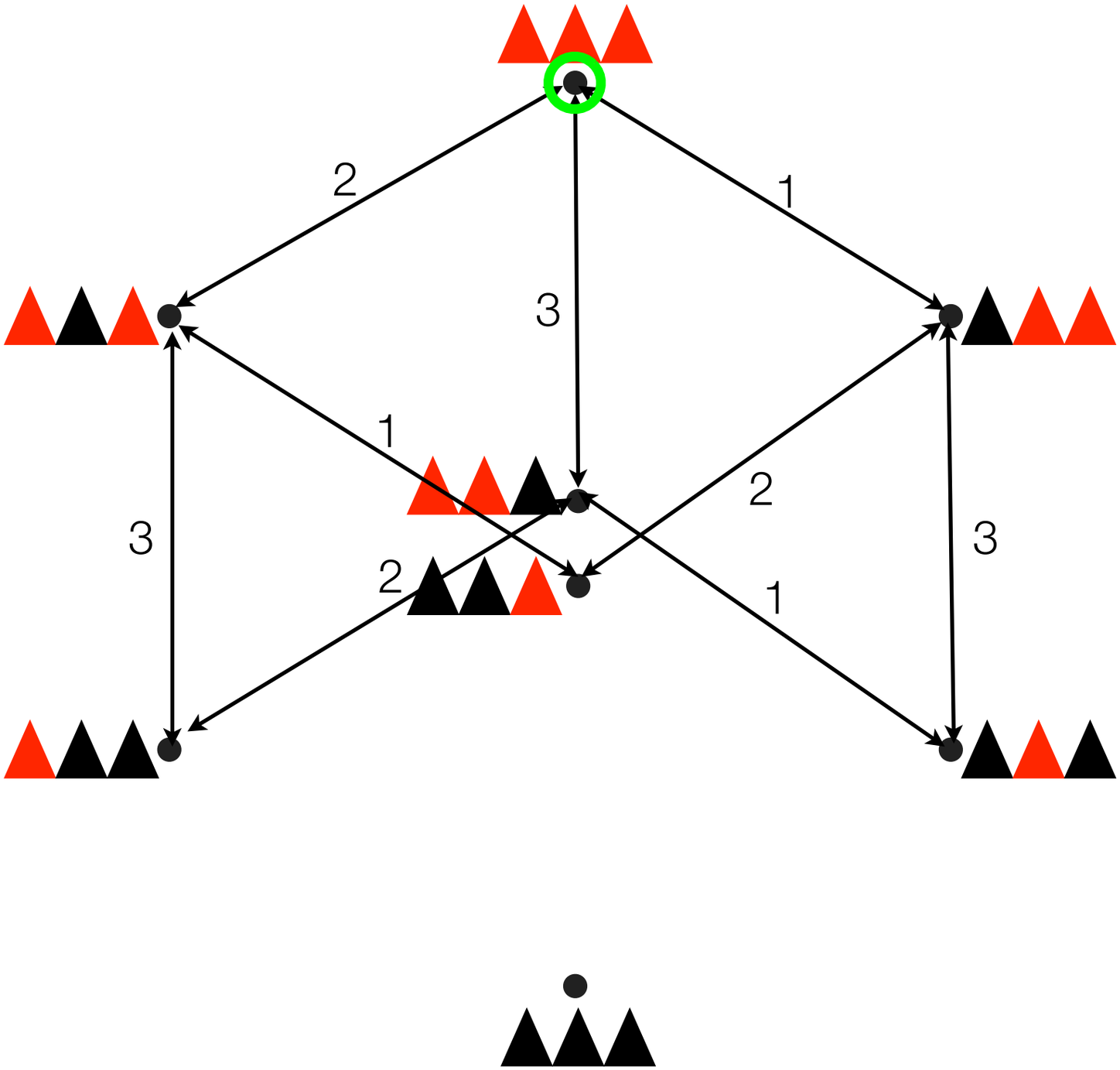
At least one hat is red

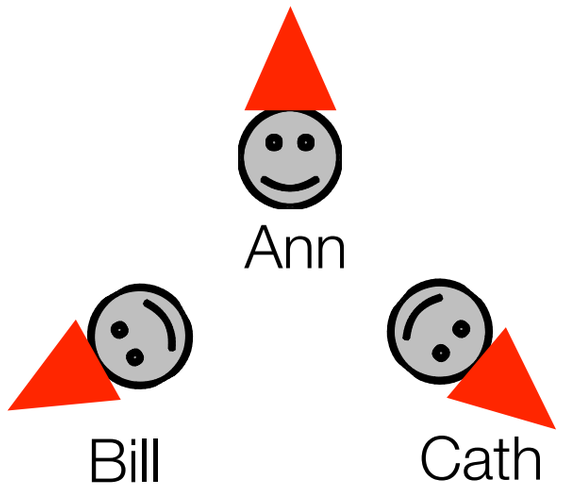
p



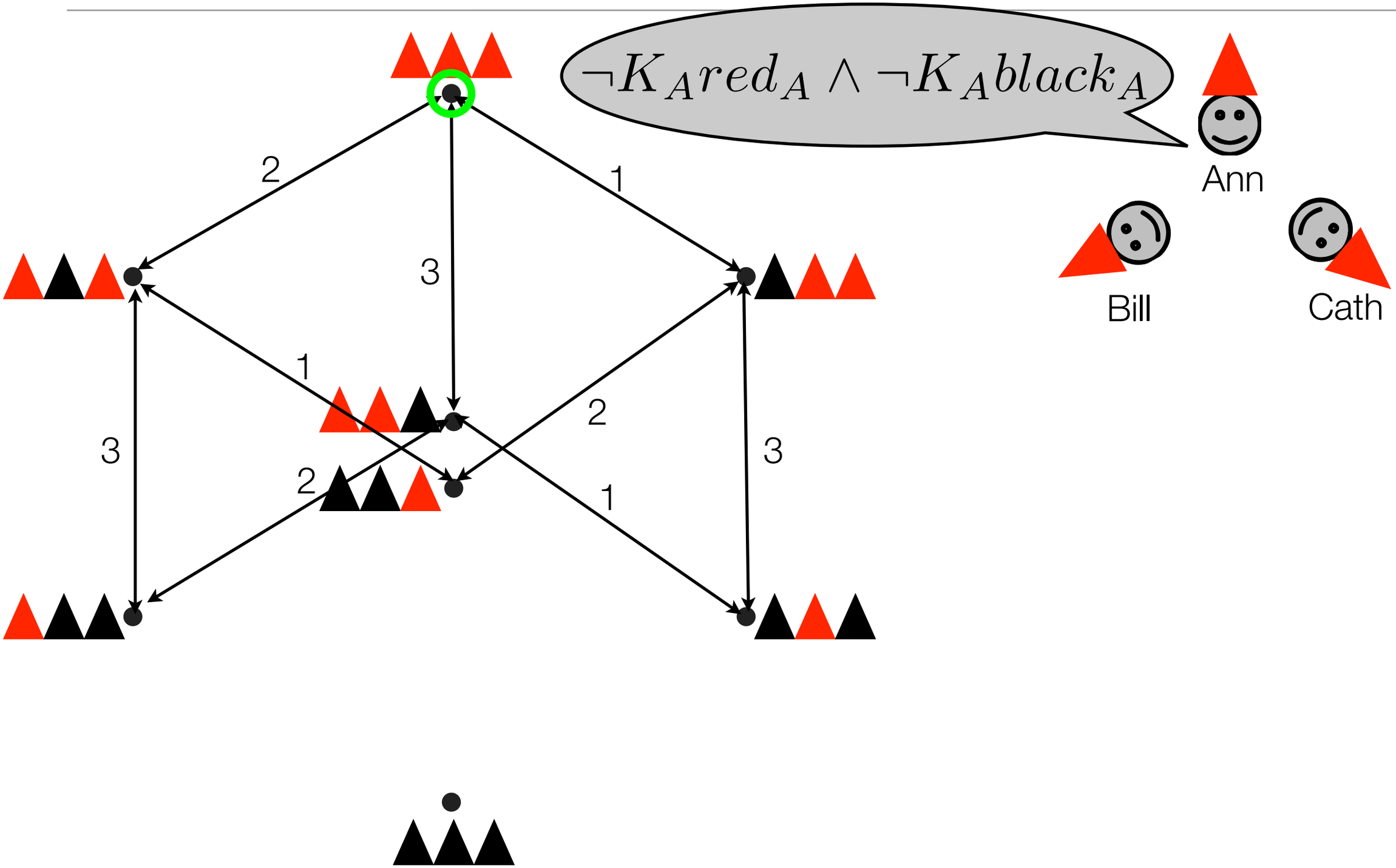
$\neg p$

3 Chinese students: after the announcement

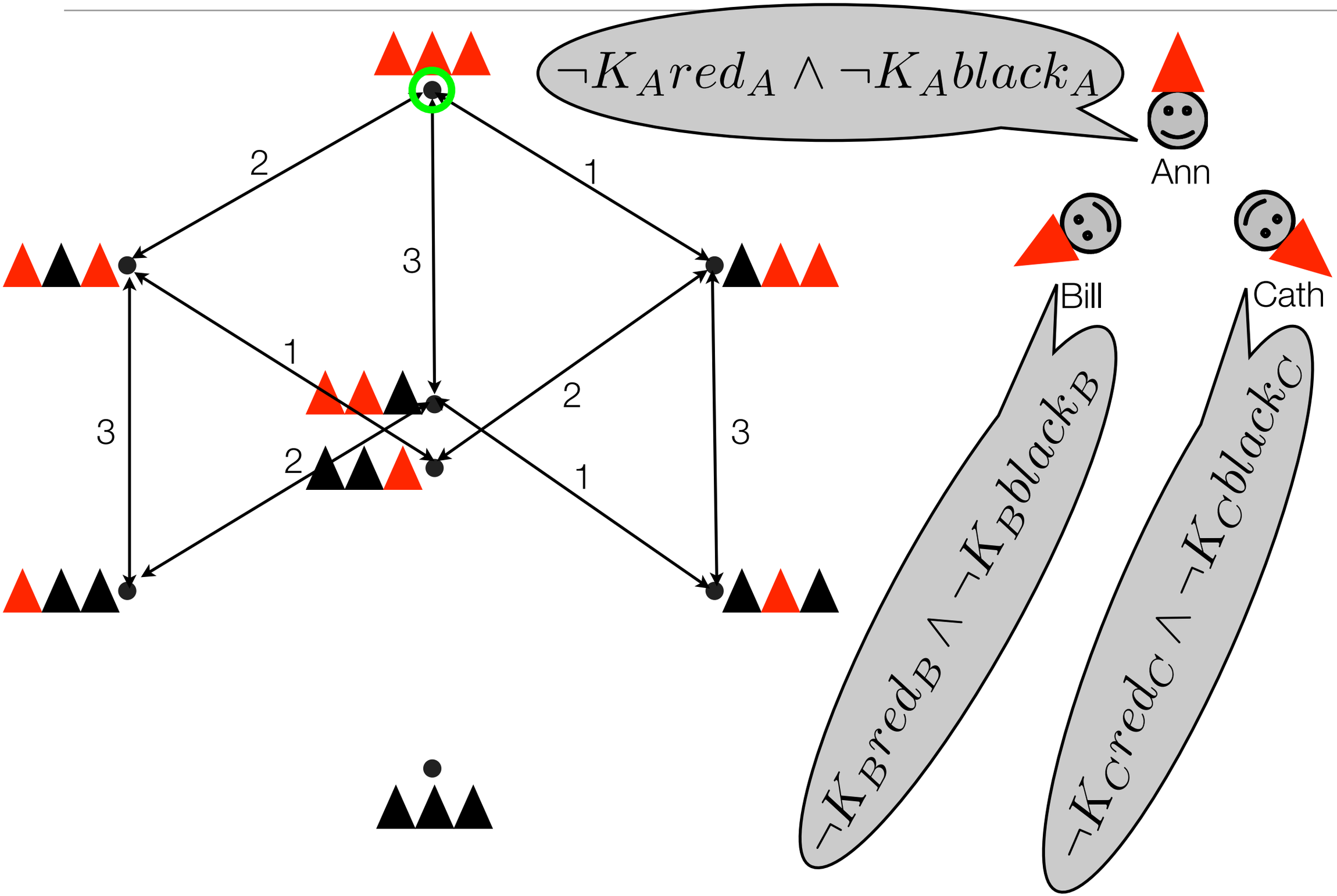




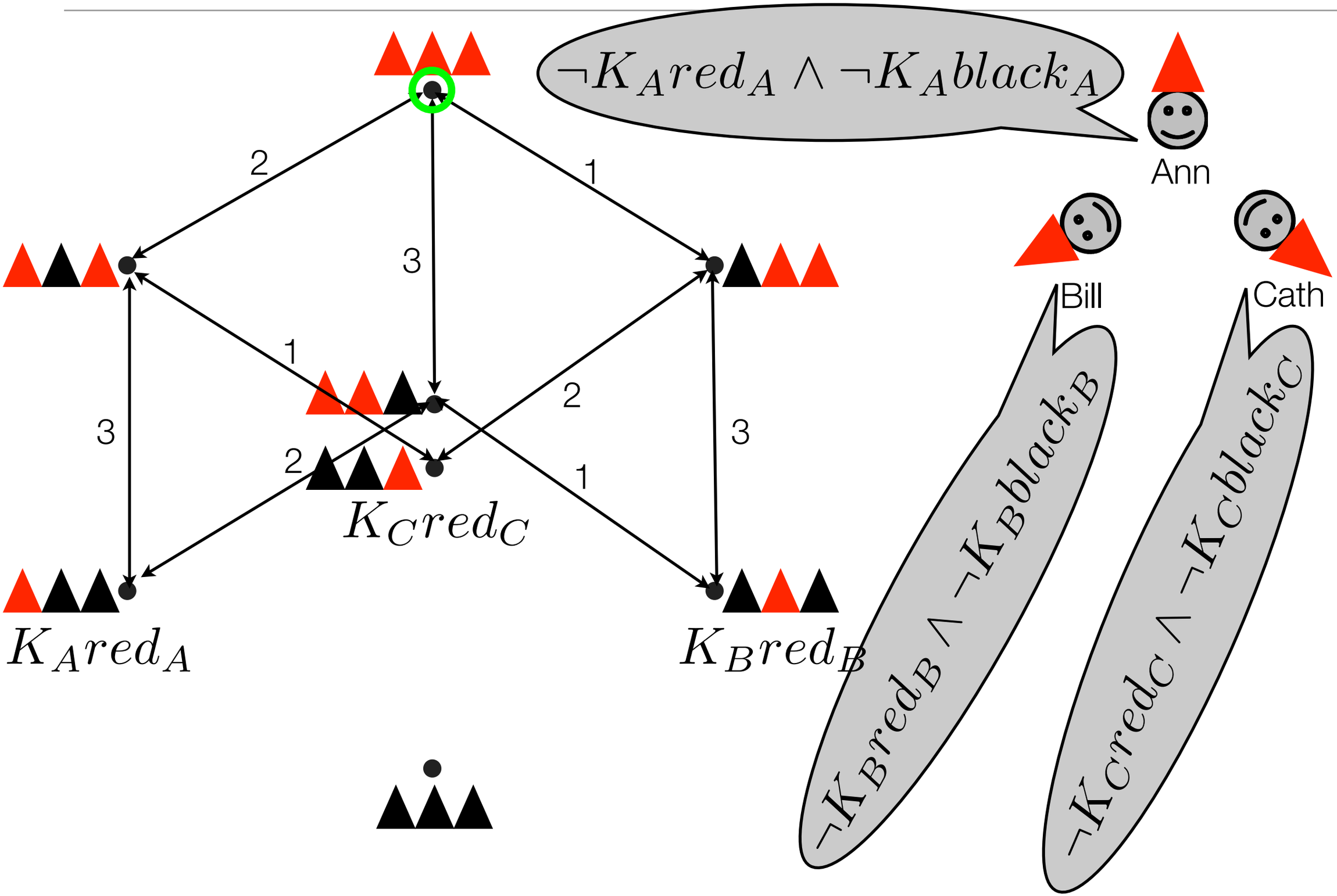
3 Chinese students: after the announcement



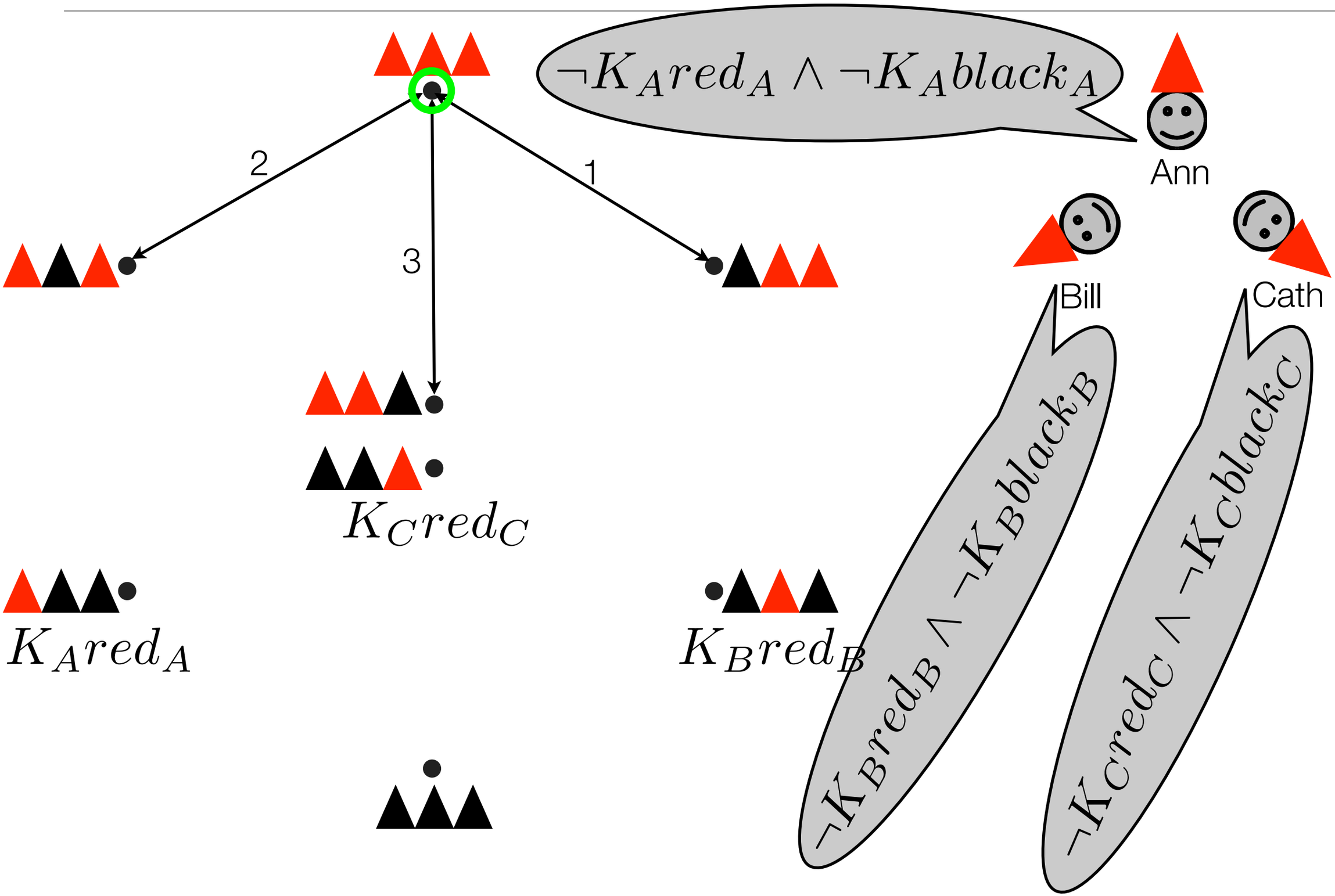
3 Chinese students: after the announcement



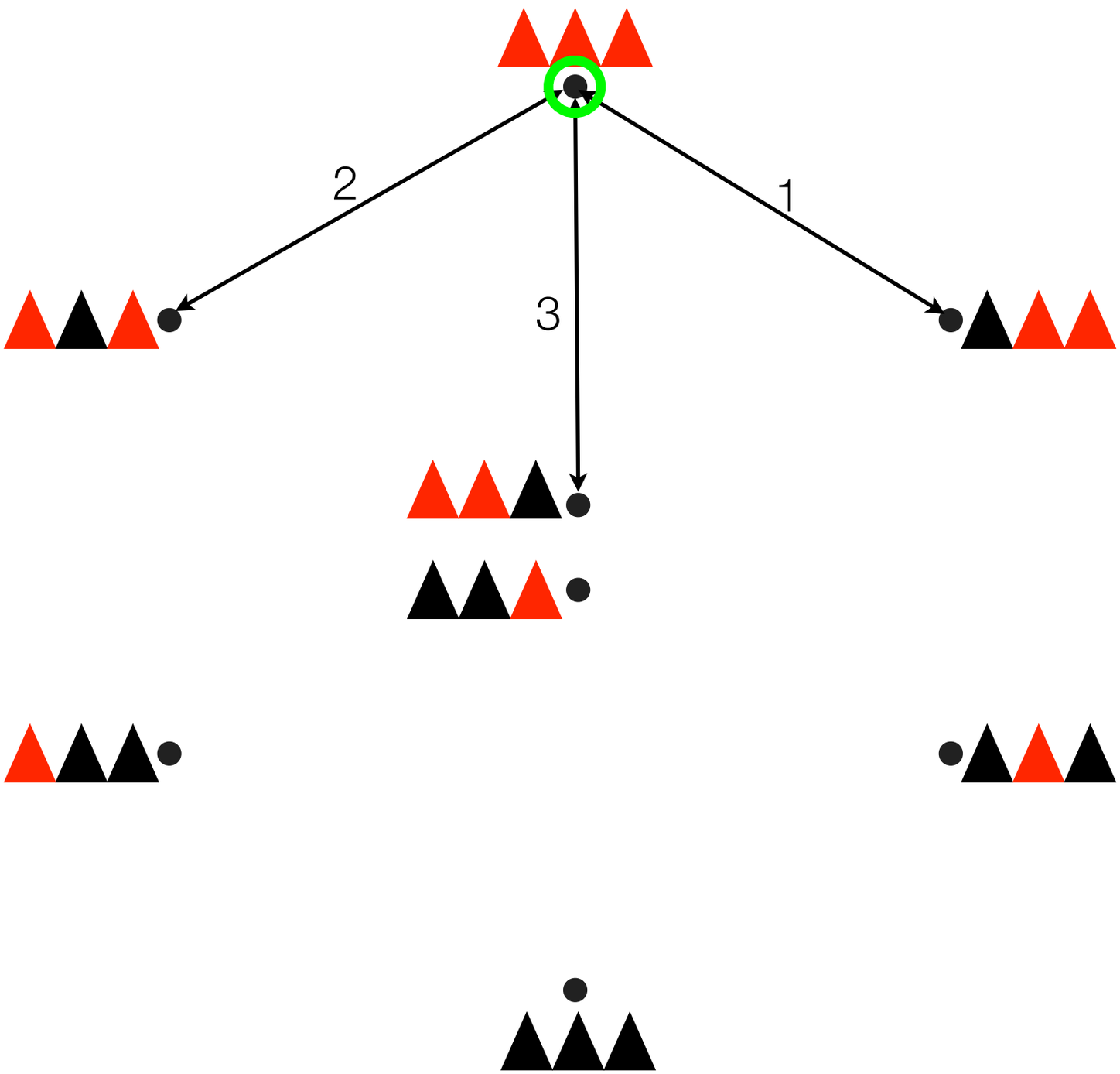
3 Chinese students: after the announcement

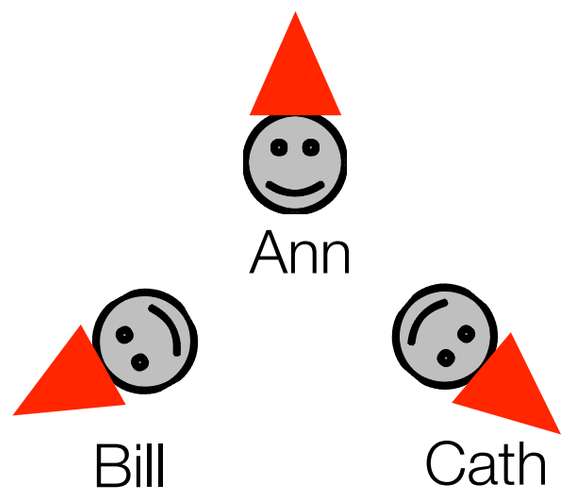


3 Chinese students: after the announcement

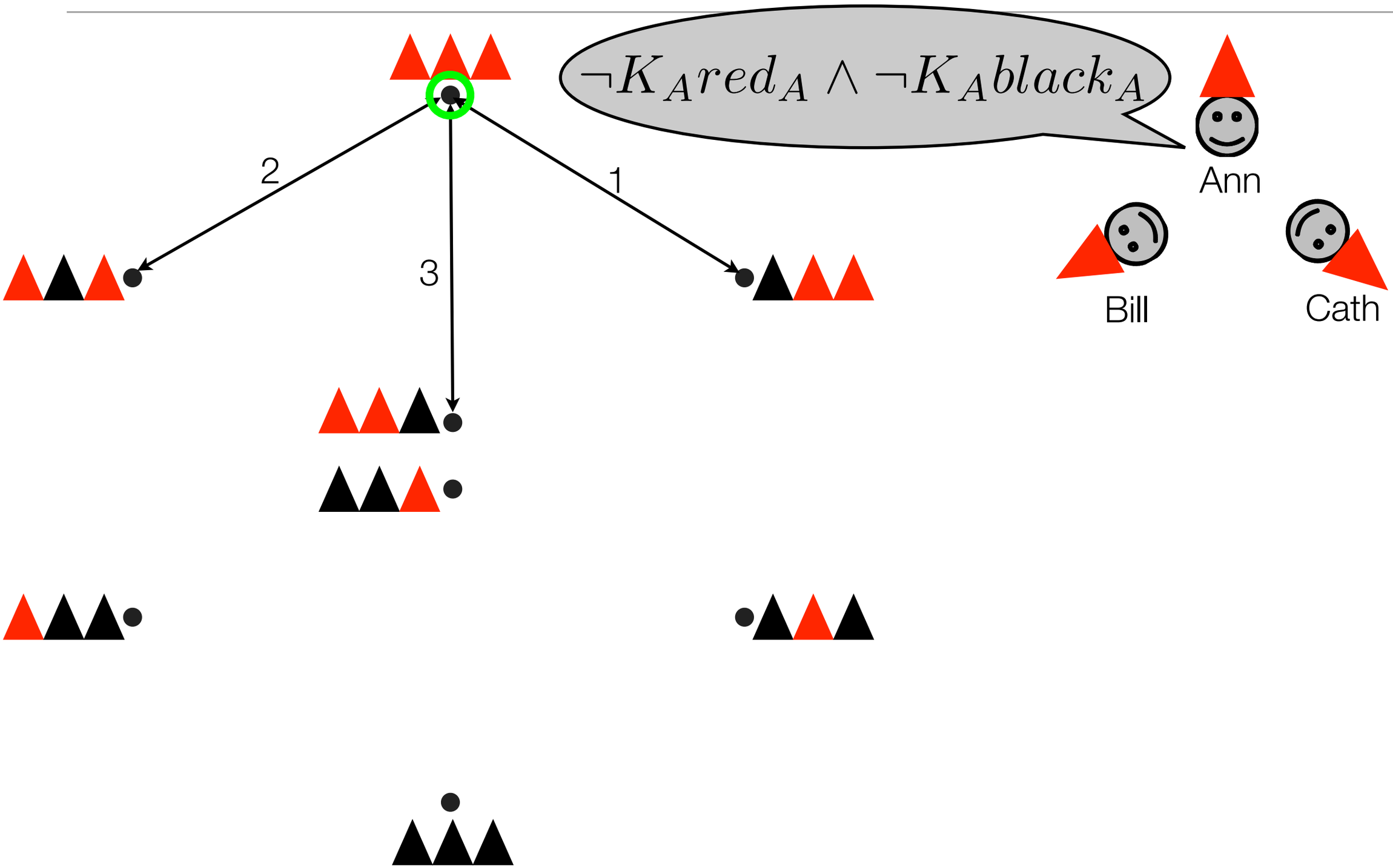


3 Chinese students: after answers to the first question

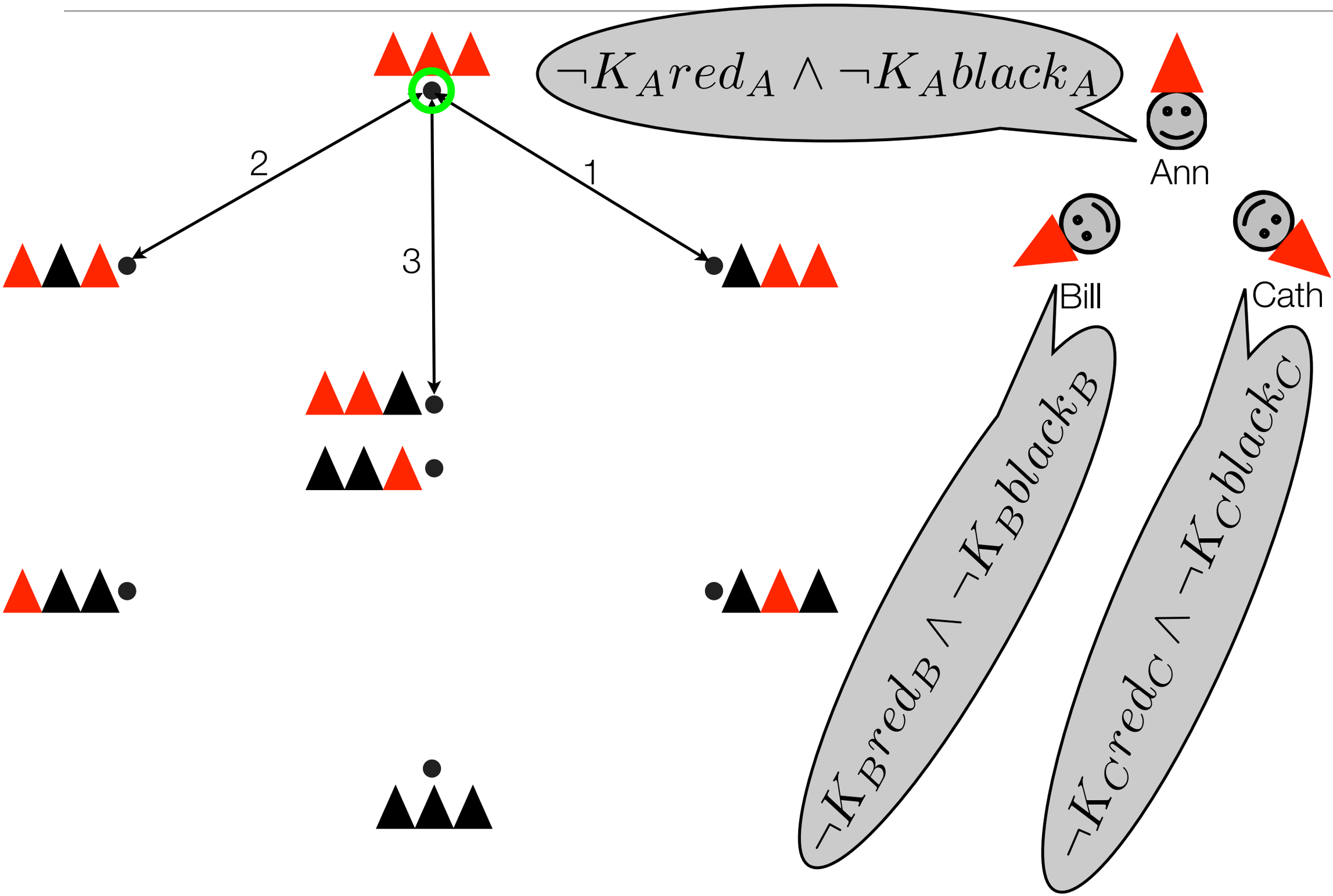




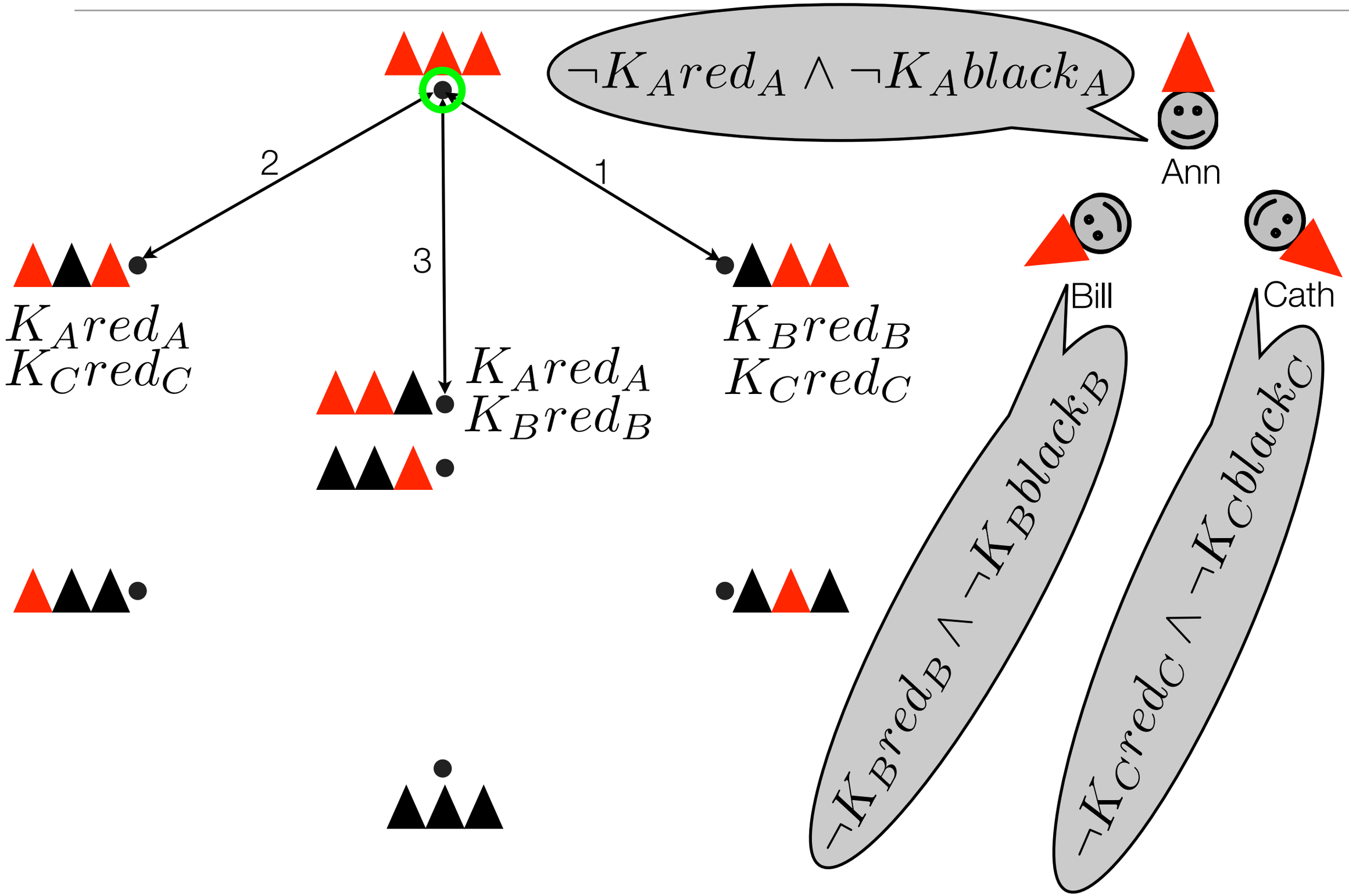
3 Chinese students: after answers to the first question



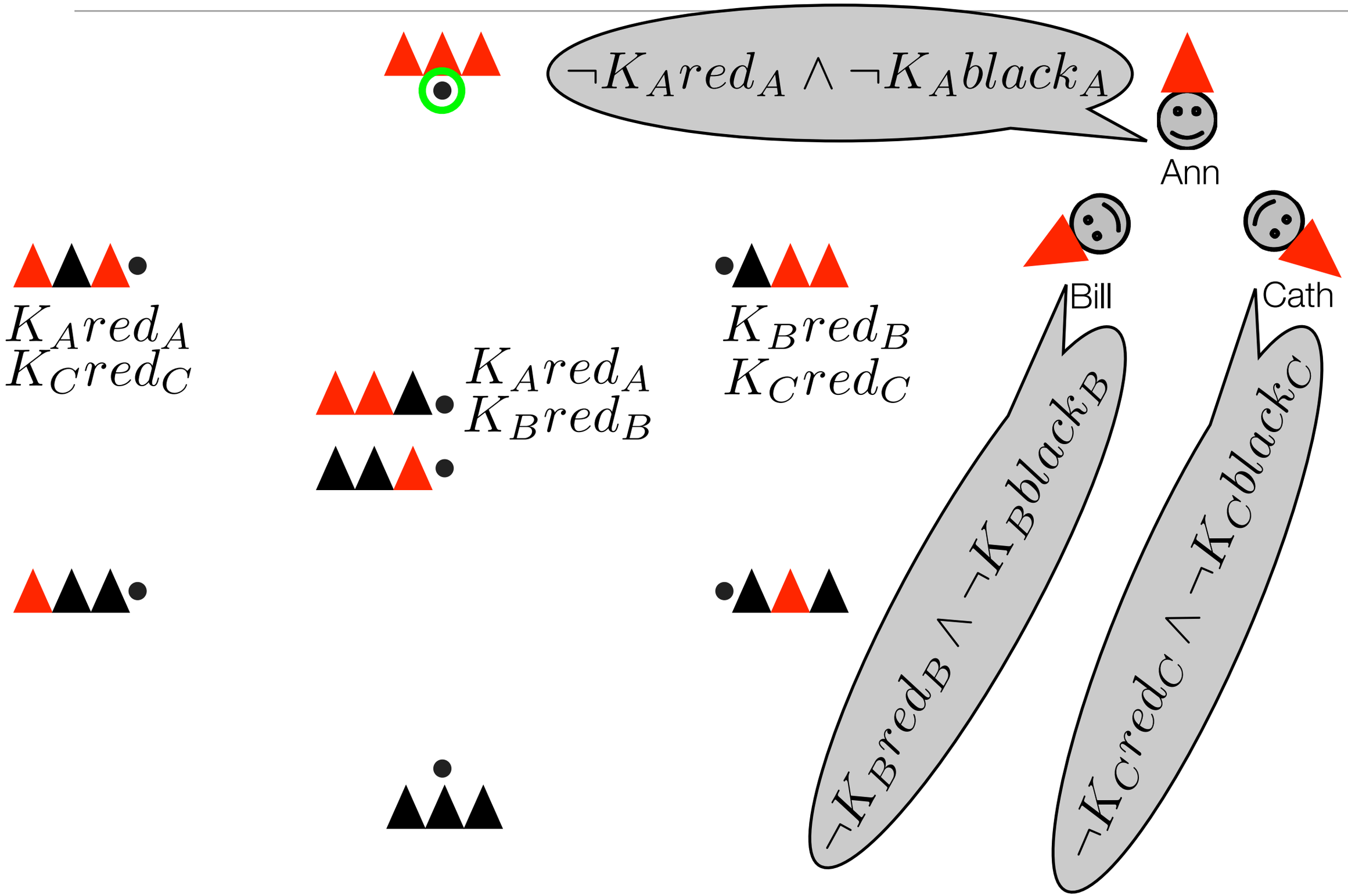
3 Chinese students: after answers to the first question



3 Chinese students: after answers to the first question

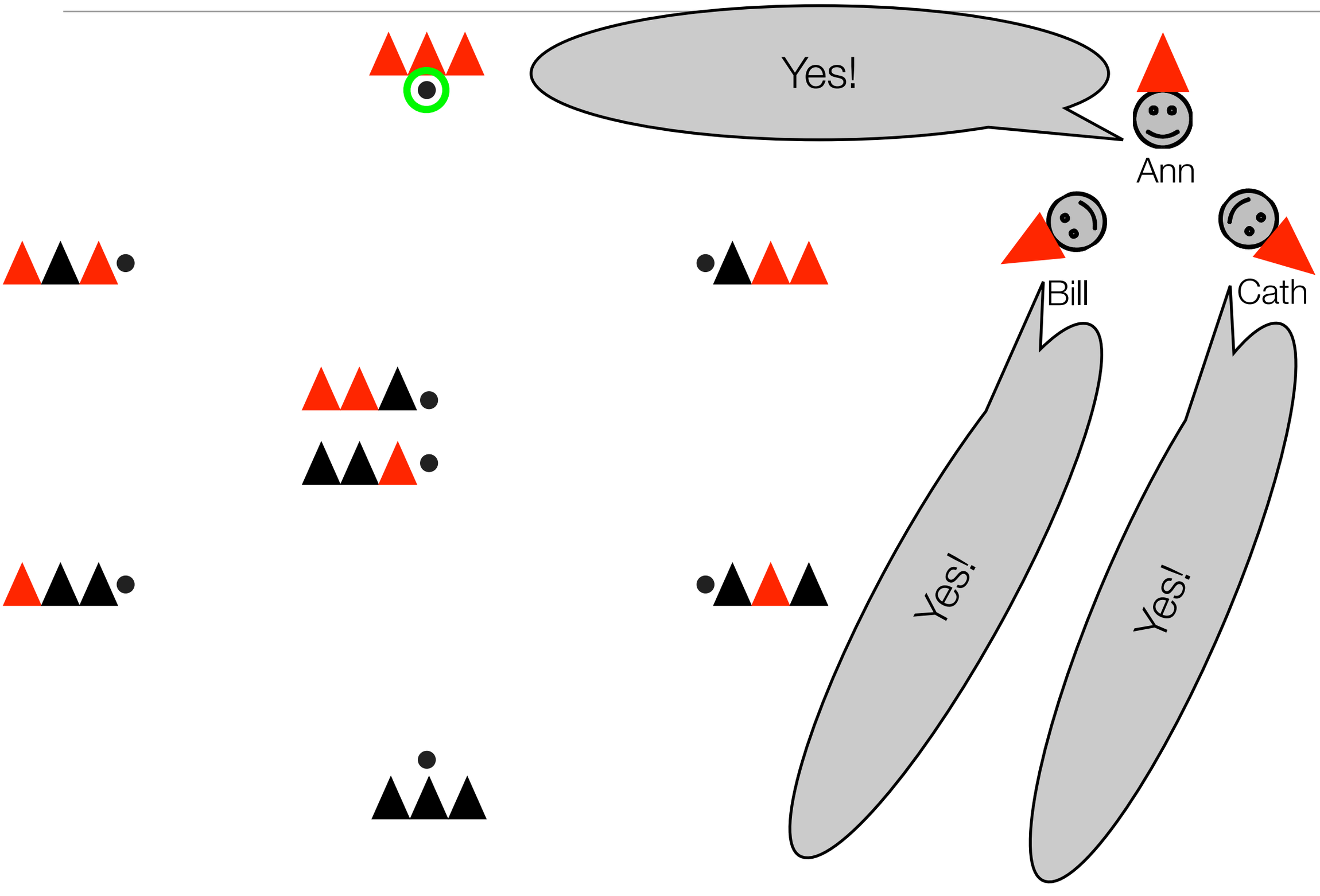


3 Chinese students: after answers to the first question

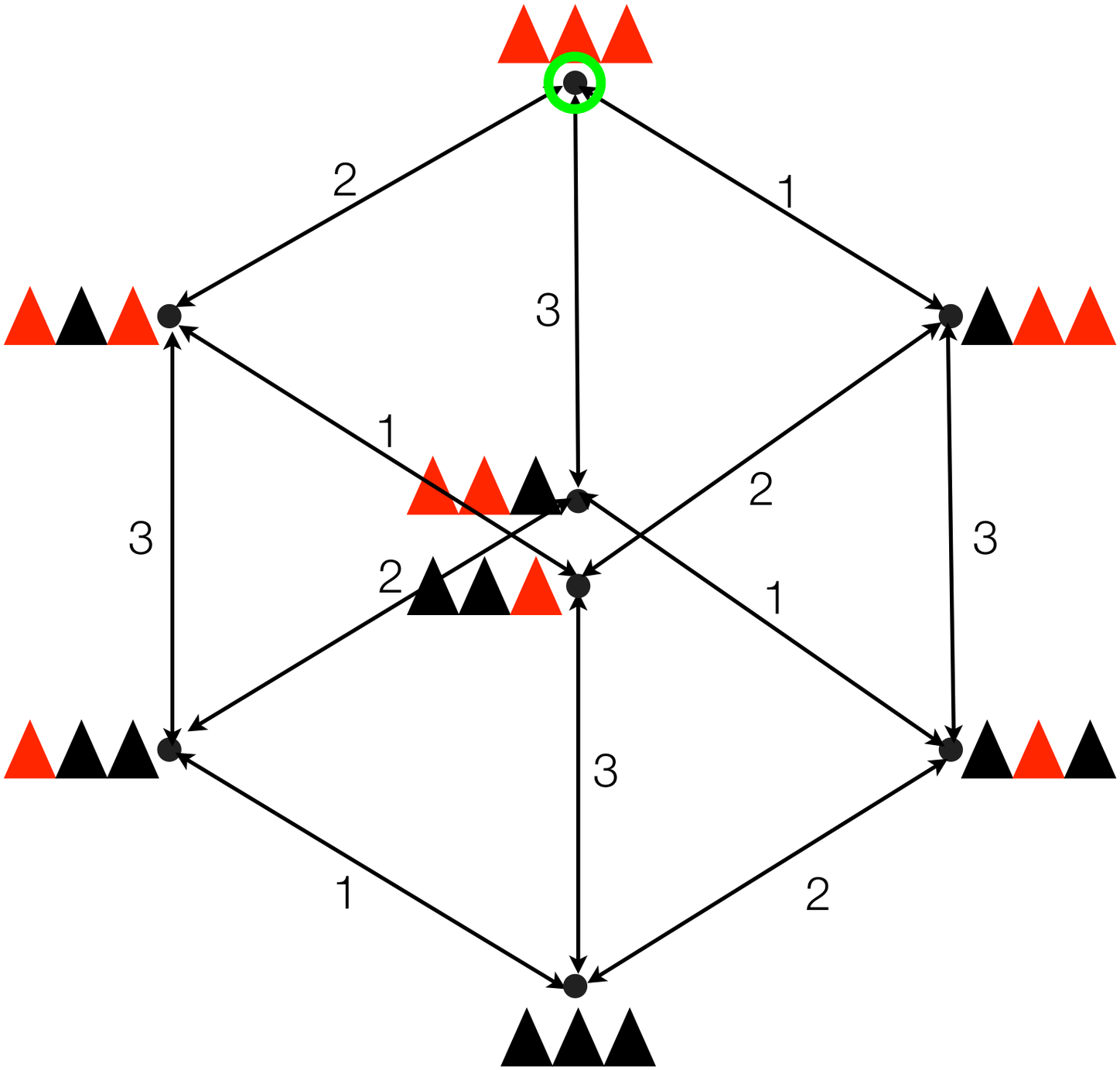


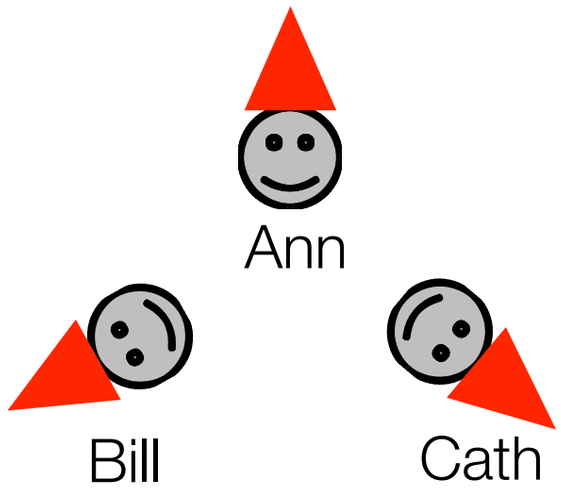
3 Chinese students: after answers to the second question

3 Chinese students: after answers to the second question

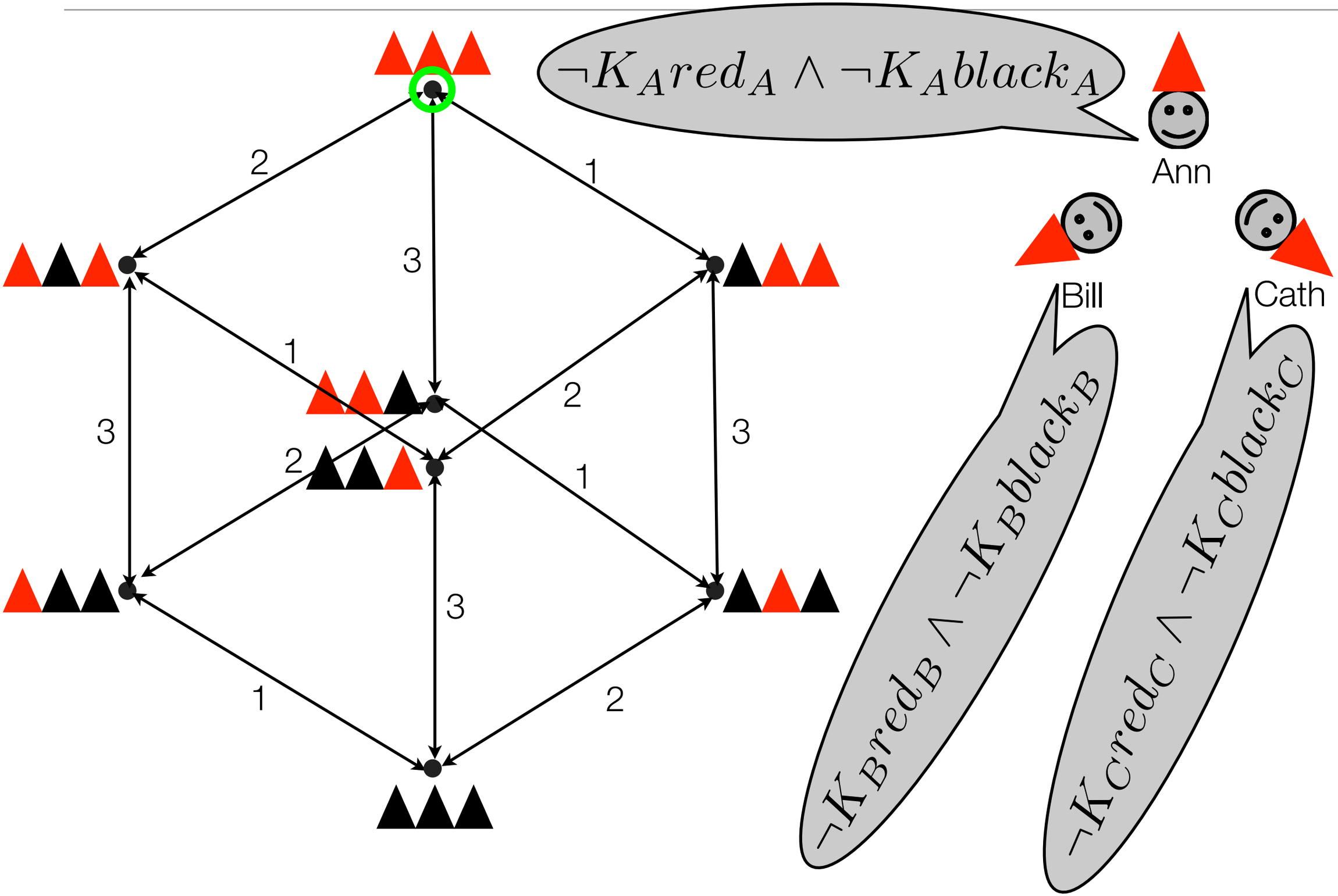


3 Chinese students: without the announcement

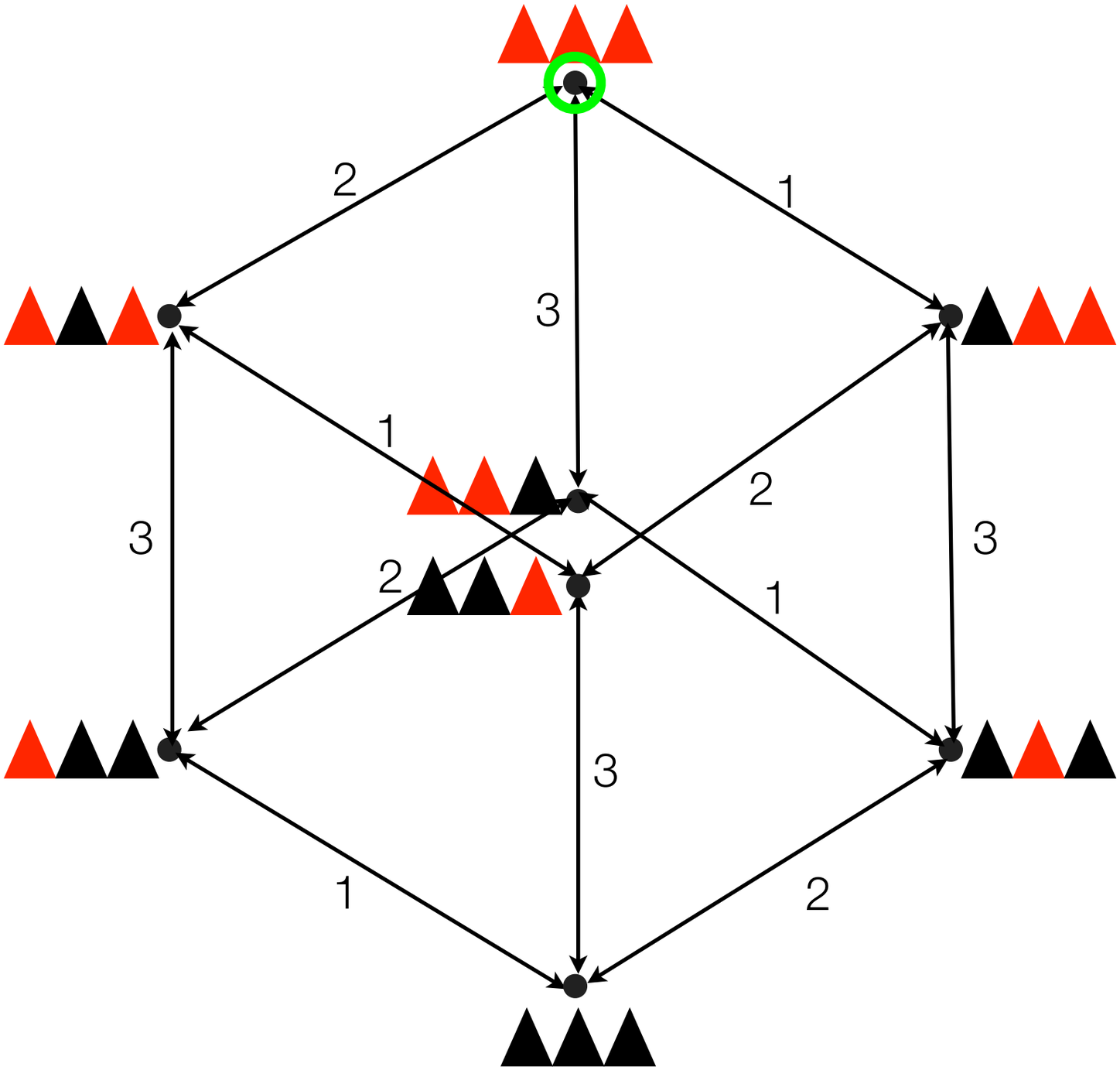


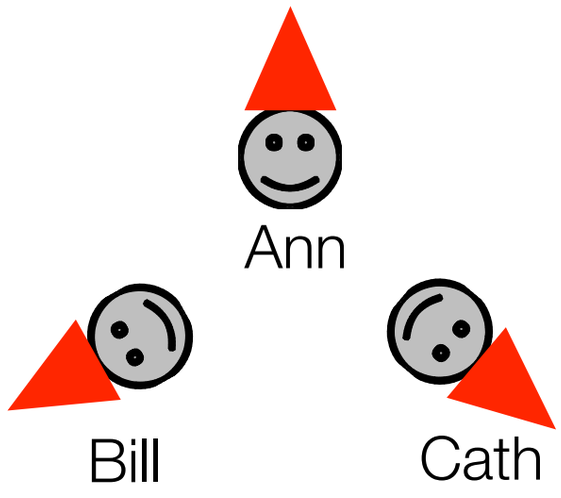


3 Chinese students: without the announcement

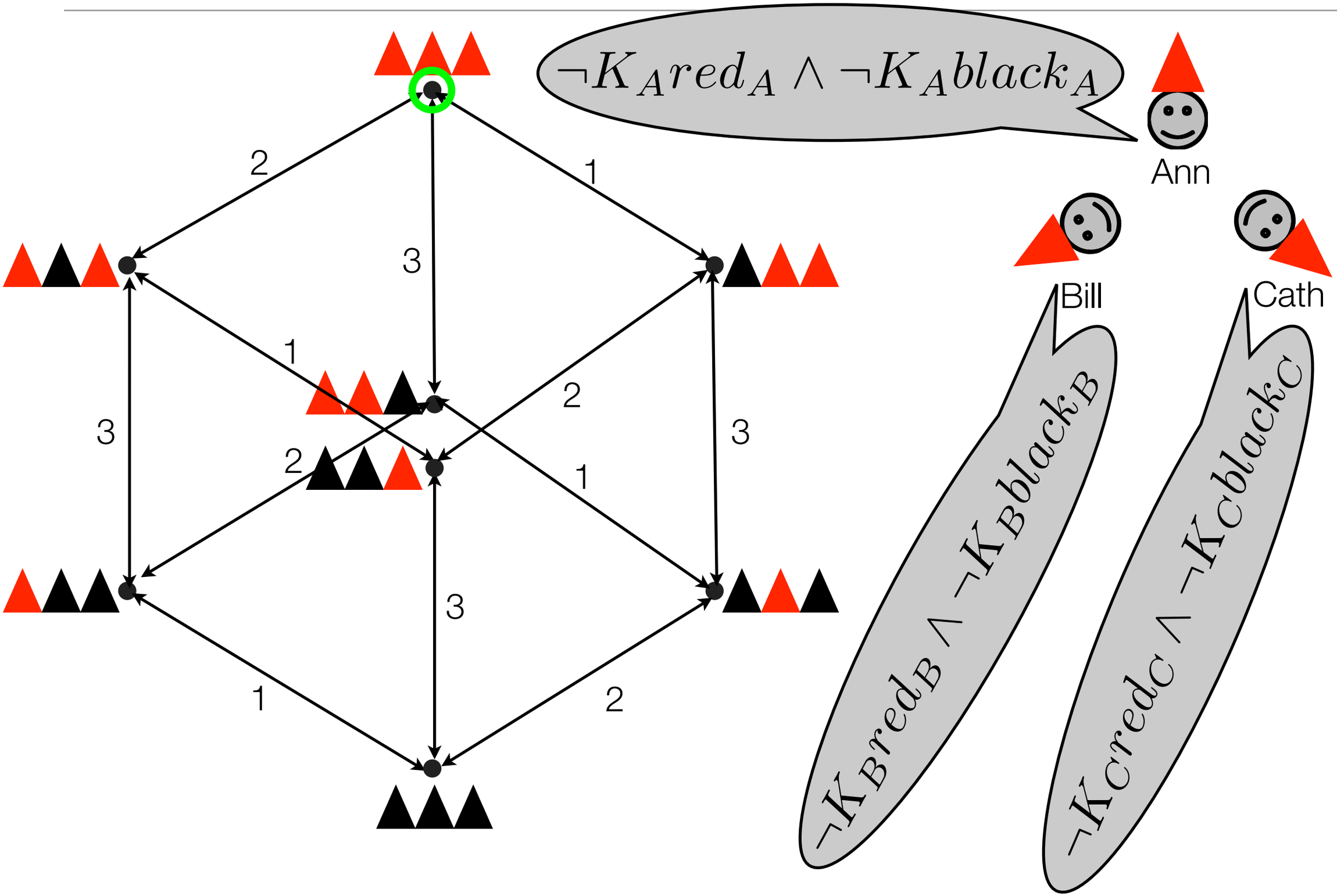


3 Chinese students: without the announcement, after first question

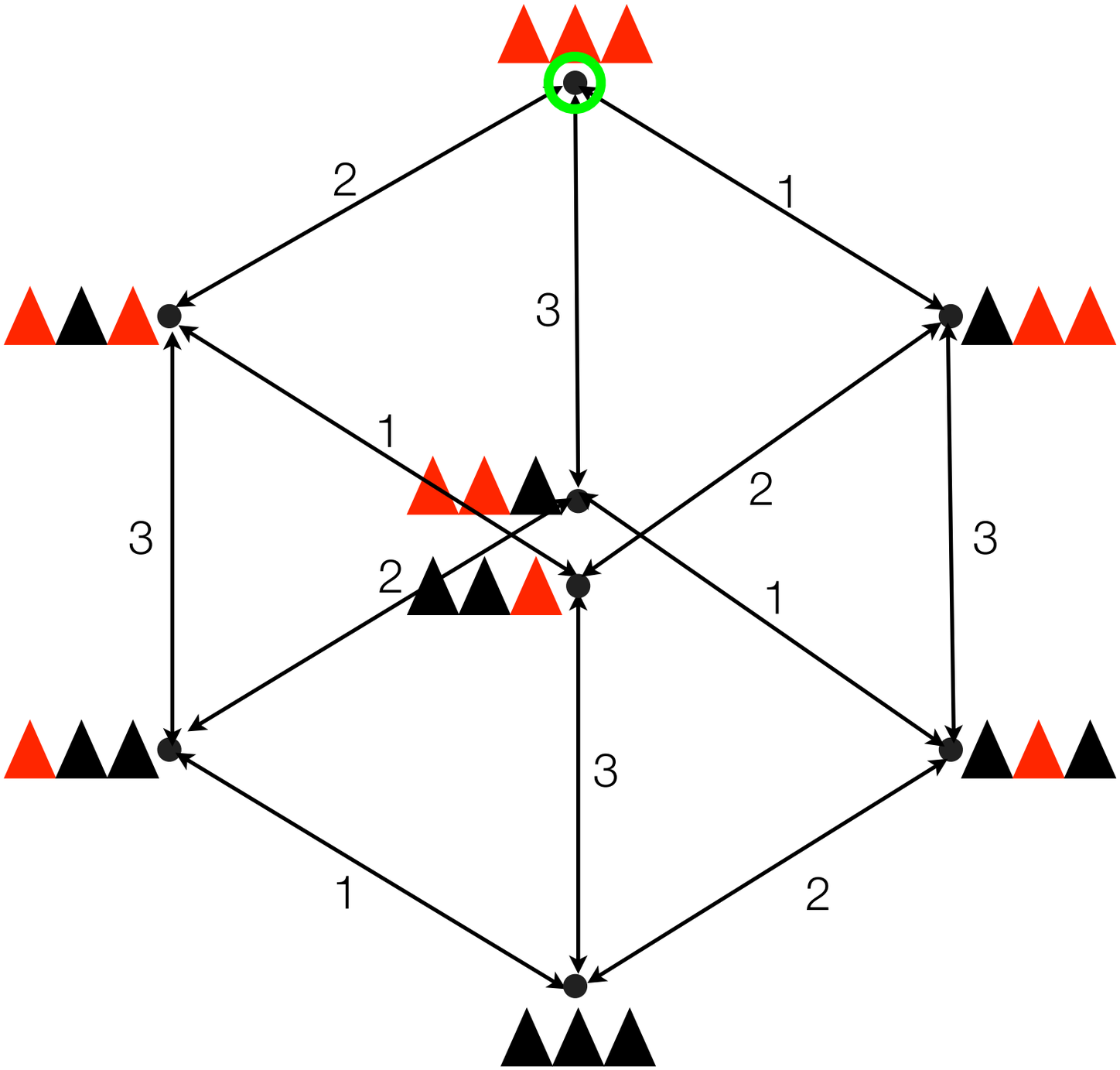


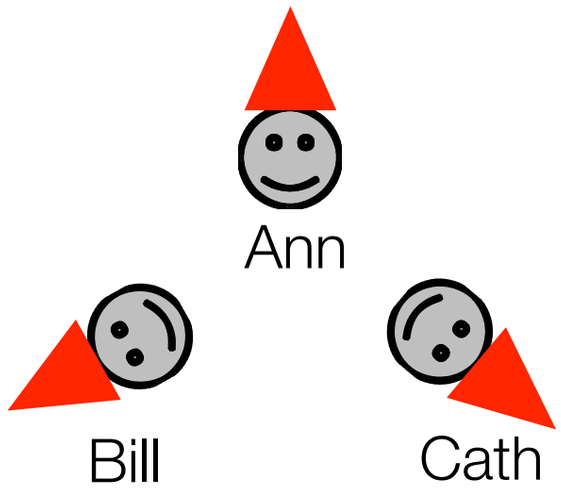


3 Chinese students: without the announcement, after first question

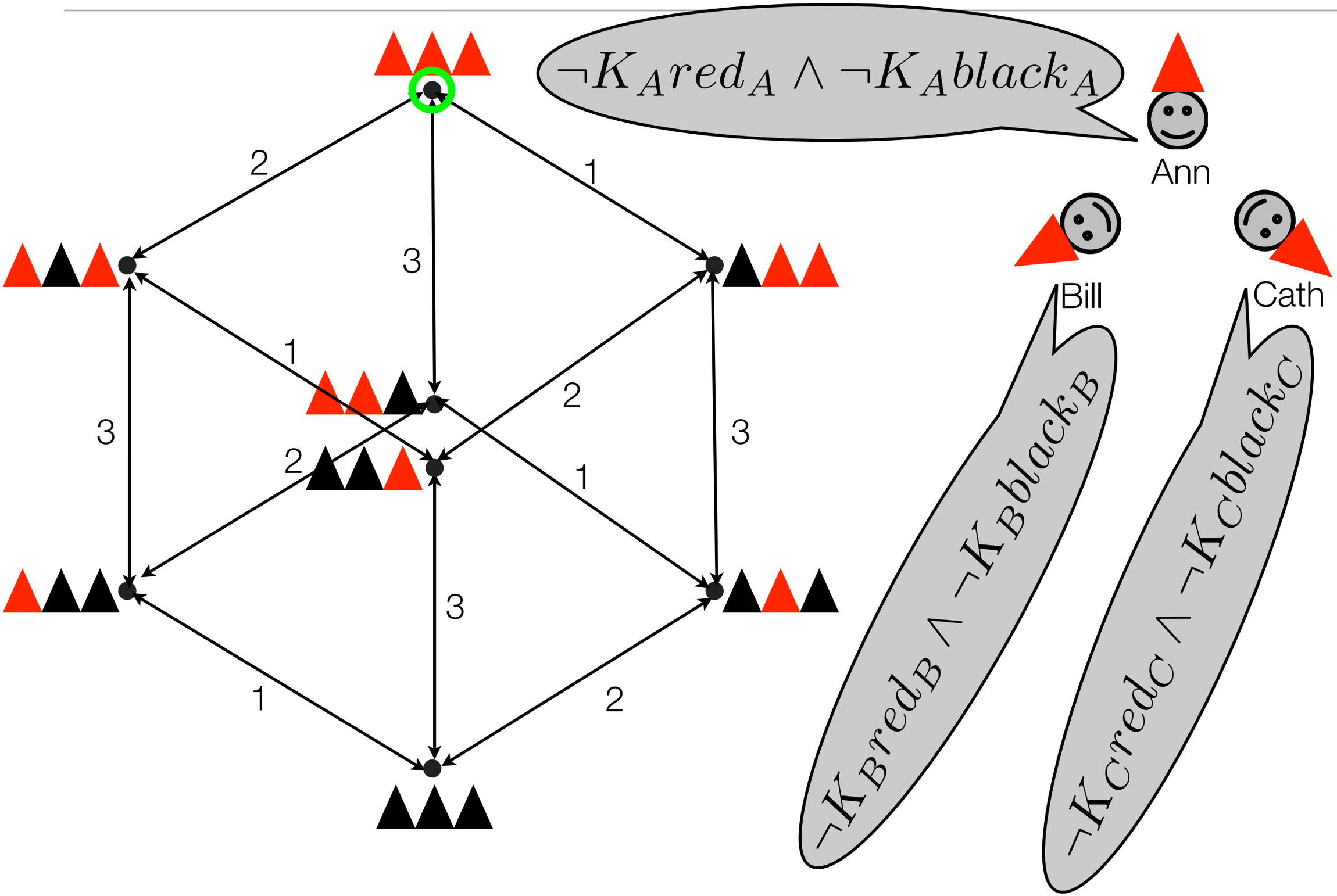


3 Chinese students: without the announcement, after second question

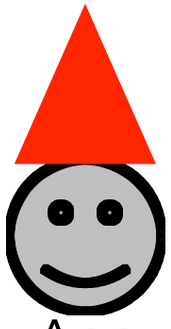


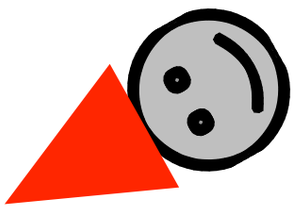


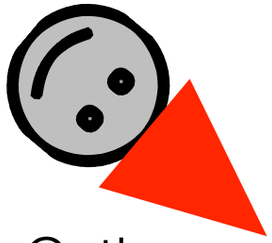
3 Chinese students: without the announcement, after second question



3 Chinese students: crucial assumptions


Ann


Bill


Cath

- All the hats are red
- You can only see the others' hats, not your own

It is well known that Chinese students are:

- extremely intelligent
- truthful
- perceptive

Thank you!

谢谢！

Challenge

A says to S and P: I have chosen two integers x, y such that $1 < x < y$ and $x + y \leq 100$. In a moment, I will inform S only of $s = x + y$, and P only of $p = xy$. These announcements remain private. You are required to determine the pair (x, y).

He acts as said. The following conversation now takes place:

i. P says: "I do not know it."

ii. S says: "I knew you didn't."

iii. P says: "I now know it."

iv. S says: "I now also know it."

Determine the pair (x, y).

Send solution to: thomas.agotnes@infomedia.uib.no