# Epistemic puzzles


## Hans van Ditmarsch, University of Otago, New Zealand

hans@cs.otago.ac.nz
http://www.cs.otago.ac.nz/staffpriv/hans/


## June 2006

# Epistemic puzzles

- consecutive numbers
- *public announcement logic*
- ~~muddy children~~ — so don't worry!
- what is my number?
- sum and product
- russian cards
- darkness at noon

# Consecutive numbers
# (also known as the 'conway paradox')

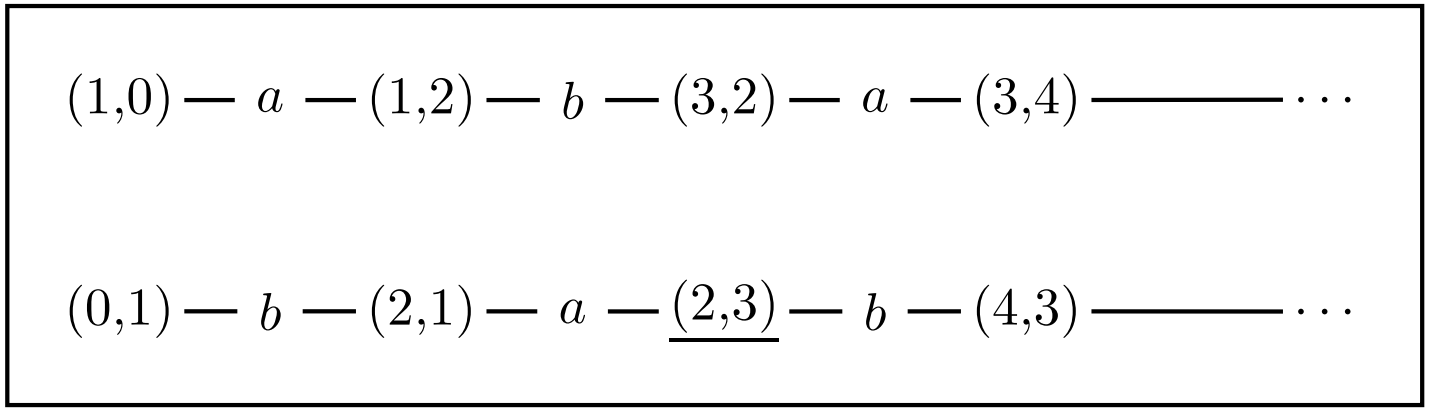
*Anne and Bill are each going to be told a natural number. Their numbers will be one apart. And they are aware of this scenario. The numbers are now being whispered in their respective ears. Suppose Anne is told 2 and Bill is told 3.*

*The following truthful conversation between Anne and Bill now takes place:*

- *Anne: "I do not know your number."*
- *Bill: "I do not know your number."*
- *Anne: "I know your number."*
- *Bill: "I know your number."*

*Explain why is this possible.*

# Consecutive numbers — representing uncertainties




- Anne knows that her number is 2.
- Bill knows that Anne's number is 2 or 4.
- Anne and Bill commonly know that Bill's number is odd.
- ...

# Consecutive numbers — successive announcements

$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \text{———} \cdots$

$(0,1) - b - (2,1) - a - \underline{(2,3)} - b - (4,3) \text{———} \cdots$

- Anne: "I do not know your number." ??

# Consecutive numbers — successive announcements

$$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \text{———} \cdots$$

$$(0,1) - b - (2,1) - a - \underline{(2,3)} - b - (4,3) \text{———} \cdots$$

- Anne: "I do not know your number." eliminated states

# Consecutive numbers — successive announcements

$$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \text{———} \cdots$$

$$(2,1) - a - \underline{(2,3)} - b - (4,3) \text{———} \cdots$$

- Anne: "I do not know your number."

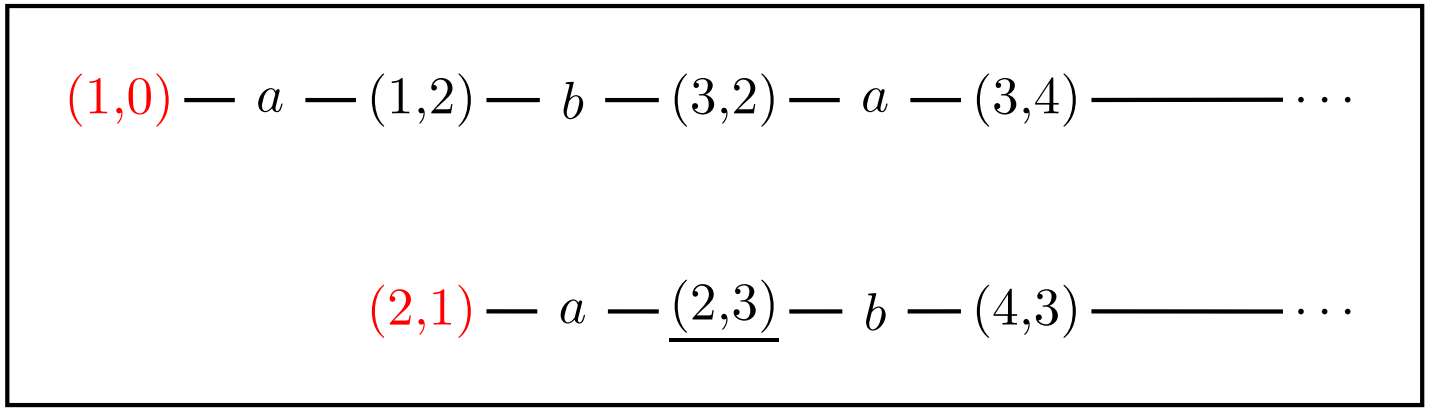
# Consecutive numbers — successive announcements

$$(1,0) - a - (1,2) - b - (3,2) - a - (3,4) \text{———} \cdots$$

$$(2,1) - a - \underline{(2,3)} - b - (4,3) \text{———} \cdots$$

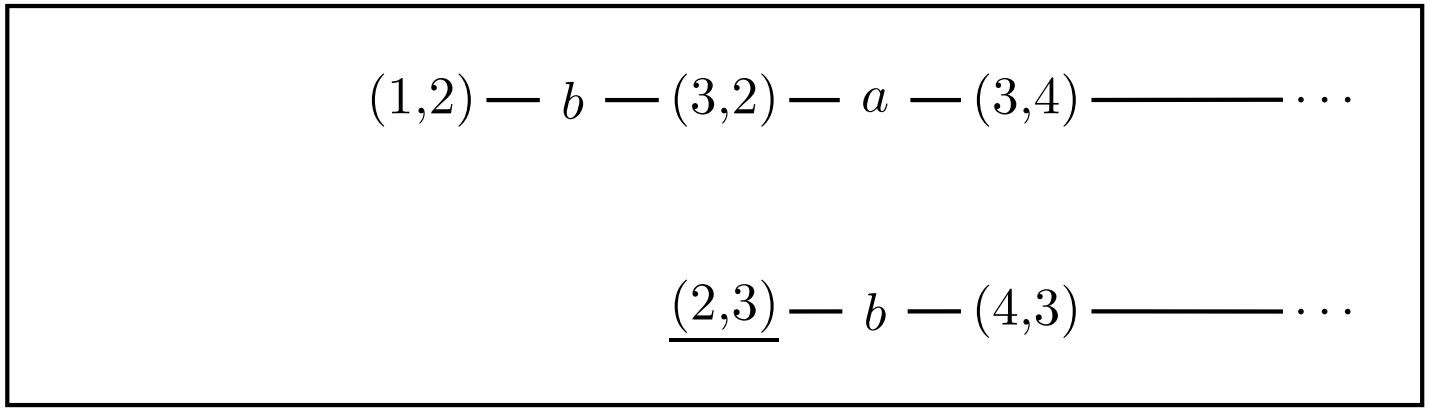
- Anne: "I do not know your number."
- Bill: "I do not know your number." ??

# Consecutive numbers — successive announcements



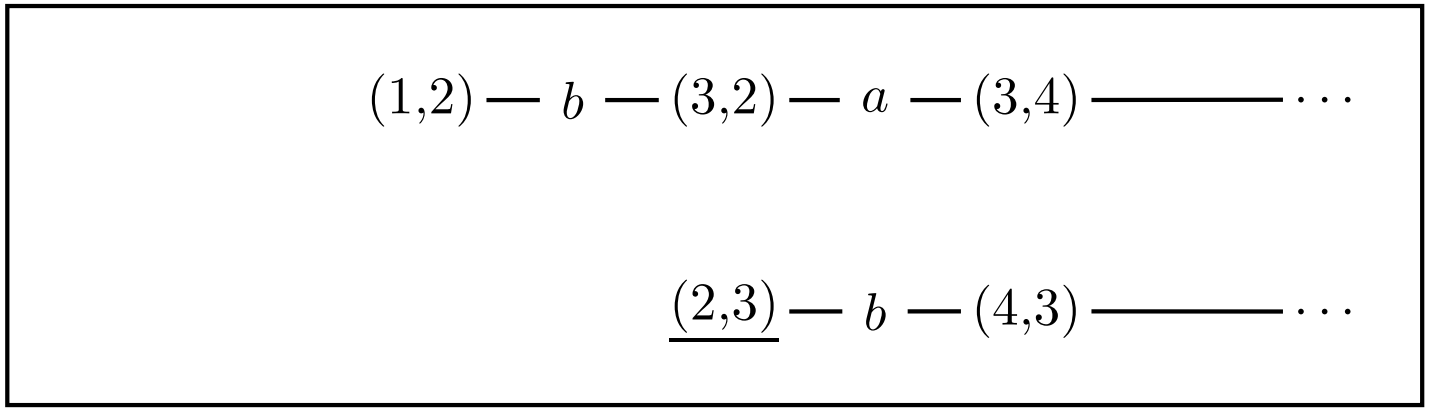

- Anne: "I do not know your number."
- Bill: "I do not know your number." eliminated states

# Consecutive numbers — successive announcements



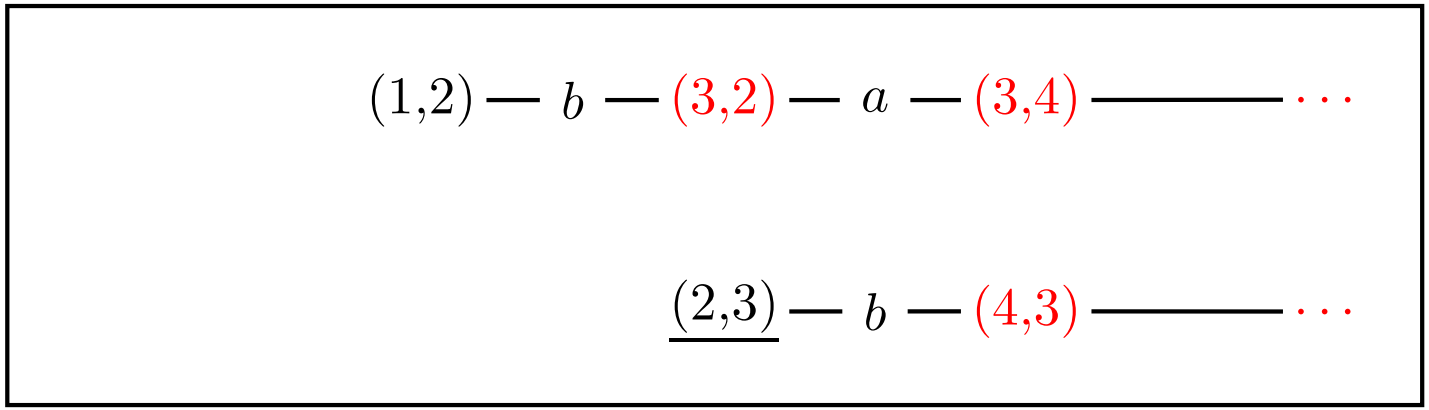

- Anne: "I do not know your number."
- Bill: "I do not know your number."

# Consecutive numbers — successive announcements



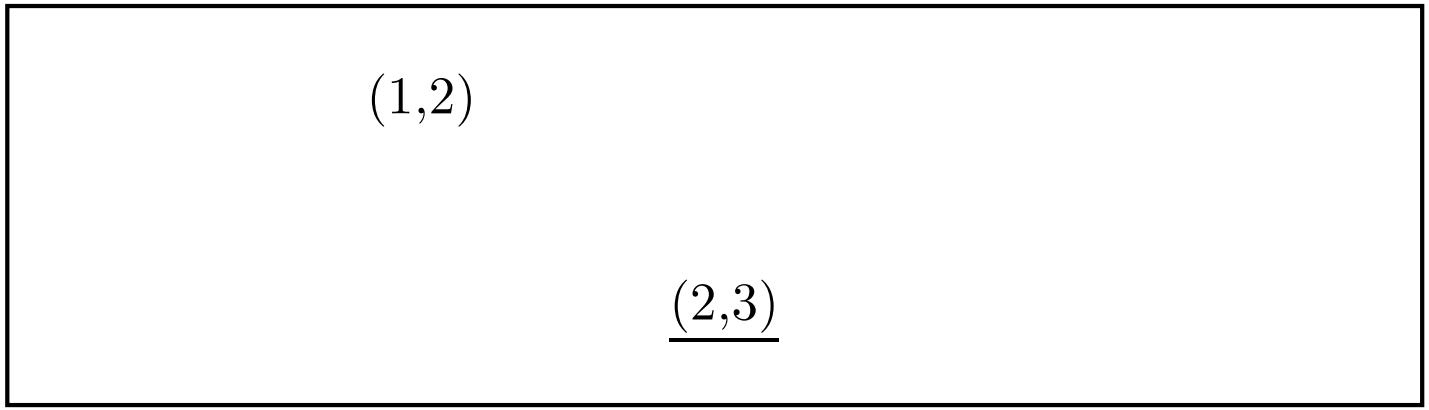

- Anne: "I do not know your number."
- Bill: "I do not know your number."
- Anne: "I know your number." ??

# Consecutive numbers — successive announcements



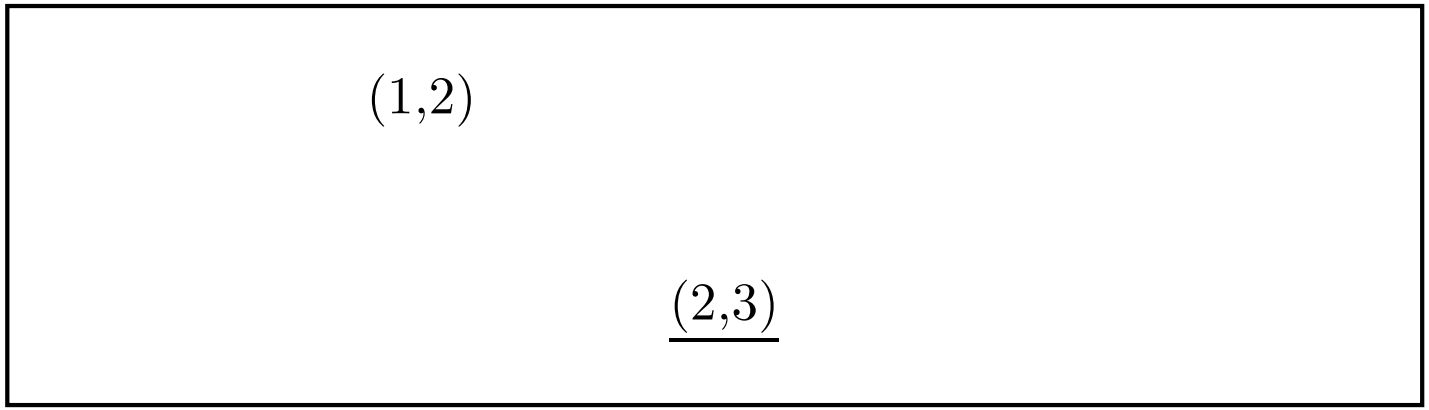

- Anne: "I do not know your number."
- Bill: "I do not know your number."
- Anne: "I know your number." eliminated states

# Consecutive numbers — successive announcements



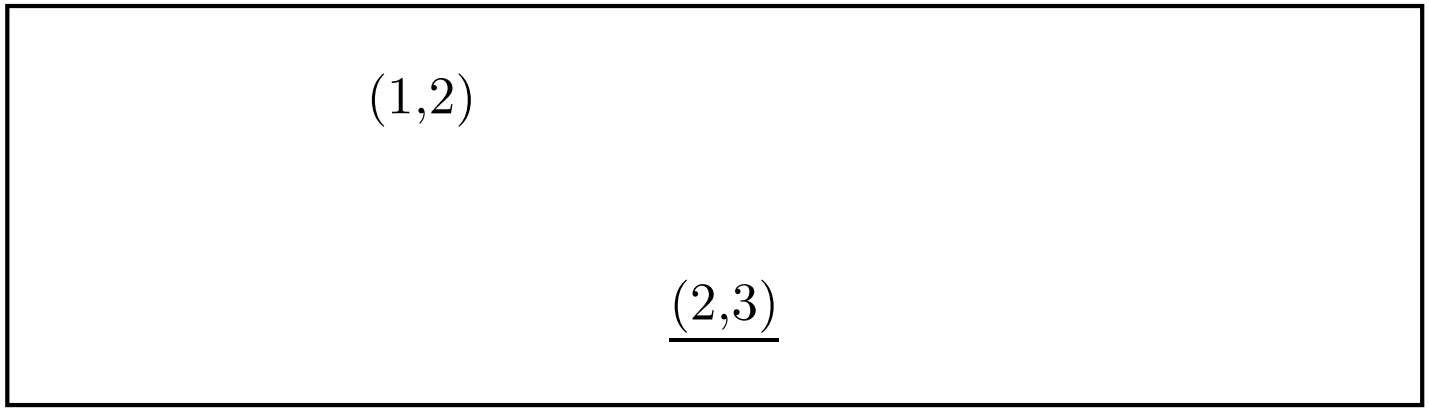

- Anne: "I do not know your number."
- Bill: "I do not know your number."
- Anne: "I know your number."

# Consecutive numbers — successive announcements



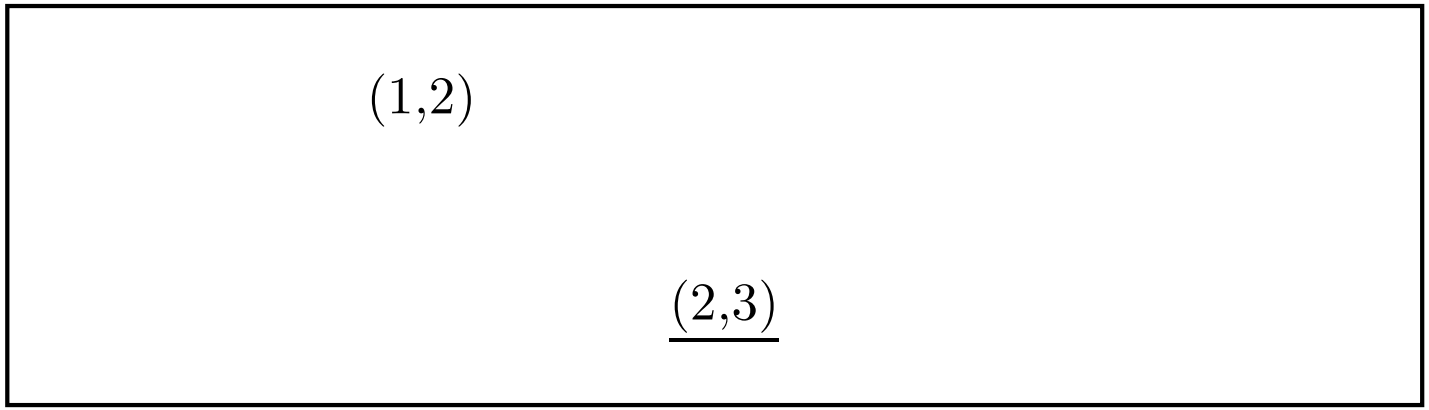

- Anne: "I do not know your number."
- Bill: "I do not know your number."
- Anne: "I know your number."
- Bill: "I know your number." ??

# Consecutive numbers — successive announcements




- Anne: "I do not know your number."
- Bill: "I do not know your number."
- Anne: "I know your number."
- Bill: "I know your number." already common knowledge

# Consecutive numbers — successive announcements




- Anne: "I do not know your number."
- Bill: "I do not know your number."
- Anne: "I know your number."
- Bill: "I know your number."

# Epistemic puzzles as interpreted systems

*Anne and Bill are each going to be told a natural number. Their numbers will be one apart. And they are aware of this scenario. The numbers are now being whispered in their respective ears. Suppose Anne is told 2 and Bill is told 3.*

*Anne and Bill each have a natural number on their forehead. Their numbers are one apart. They only can see the number on the other's forehead. And they are aware of this scenario. ('All the previous is common knowledge.') Suppose Anne has the number 3 and Bill has the number 2.*

**There is no difference!**

# Public Announcement Logic — structures

*epistemic model* $M = \langle S, \sim, V \rangle$:

- *domain* $S$ of (factual) *states* ('worlds')
- *accessibility* $\sim : A \to \mathcal{P}(S \times S)$
  (set of equivalence relations $\sim_a$)
- *valuation* $V : P \to \mathcal{P}(S)$
  (set of valuations $V_p \subseteq S$)

*epistemic state* $(M, s)$:

- For $s \in S$, $(M, s)$ is an *epistemic state*

# Public Announcement Logic — language

$$\varphi \quad ::= \quad p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_a\varphi \mid C_B\varphi \mid [\varphi]\varphi$$

atoms / negation / conjunction / knowledge / common knowl. / announcement

# Public Announcement Logic — semantics

$$
\begin{array}{lll}
M, s \models p & \text{iff} & s \in V_p \\
M, s \models \varphi \wedge \psi & \text{iff} & M, s \models \varphi \text{ and } M, s \models \psi \\
M, s \models \neg\varphi & \text{iff} & M, s \not\models \varphi \\
M, s \models K_a\varphi & \text{iff} & \text{for all } t : s \sim_a t \text{ implies } M, t \models \varphi \\
M, s \models C_B\varphi & \text{iff} & \text{for all } t : s \sim_B t \text{ implies } M, t \models \varphi \\
M, s \models [\varphi]\psi & \text{iff} & M, s \models \varphi \text{ implies } M|\varphi, s \models \psi
\end{array}
$$

*Some details — we will learn by examples*

- $\sim_B$ is transitive reflexive closure of the union of all $\sim_a$.
- $M|\varphi$ is the restriction of epistemic model $M$ to the states where $\varphi$ is true.

# What is my number?

*Each of agents Anne, Bill, and Cath has a positive integer on its forehead. They can only see the foreheads of others. One of the numbers is the sum of the other two. All the previous is common knowledge. The agents now successively make the truthful announcements:*

*1. Anne: "I do not know my number."*

*2. Bill: "I do not know my number."*

*3. Cath: "I do not know my number."*

*4. Anne: "I know my number. It is 50."*

*What are the other numbers?*

Math Horizons, November 2004. Problem 182.

**When does Anne know her number, initially?**

# What is my number?

## When does Anne know her number, initially?

When the numbers had been $(2, 1, 1)$.
Anne sees the numbers 1 and 1;
her number must be 2 or 0;
0 is excluded.

**Anne knows that her number is 2.**

# What is my number?

*Each of agents Anne, Bill, and Cath has a ~~positive integer~~* **natural number** *on its forehead. They can only see the foreheads of others. One of the numbers is the sum of the other two. All the previous is common knowledge. The agents now successively make the truthful announcements:*

*1. Anne: "I do not know my number."*

*2. Bill: "I do not know my number."*

*3. Cath: "I do not know my number."*

*4. Anne: "I know my number. It is 50."*

*What are the other numbers?*

It's no longer solvable!

**When does Anne know her number, initially?**

# What is my number? — epistemic model $\mathcal{T}$

- *domain* consisting of triples $(x, y, z)$ such that
  $x = y + z$ or $y = x + z$ or $z = x + y$
- *equivalence relations* for agents such that (for Anne)
  $(x, y, z) \sim_a (x', y', z')$ iff $y = y'$ and $z = z'$
- *valuations* of atomic propositions (facts) $x_a$ (Anne has number $x$) such that
  $(x, y, z) \in V_{x_a}$

# What is my number? — formalizing announcements

1. Anne: "I do not know my number."
2. Bill: "I do not know my number."
3. Cath: "I do not know my number."
4. Anne: "I know my number. It is 50."

1. $\neg \bigvee_x K_a x_a$
2. $\neg \bigvee_y K_b y_b$
3. $\neg \bigvee_z K_c z_c$
4. $K_a 50_a$

# What is my number? — interpreted system

*Facts, knowledge, ignorance*:

1. **facts**: one of the three numbers is the sum of the two other numbers
   *enumerate the possible valuations / state descriptions*
2. **knowledge**: an agent can see the numbers on the forehead of other agents
   *an agent knows his local state*
3. **ignorance**: an agent does not know his own number
   *an agent considers possible every global state that extends his local state*

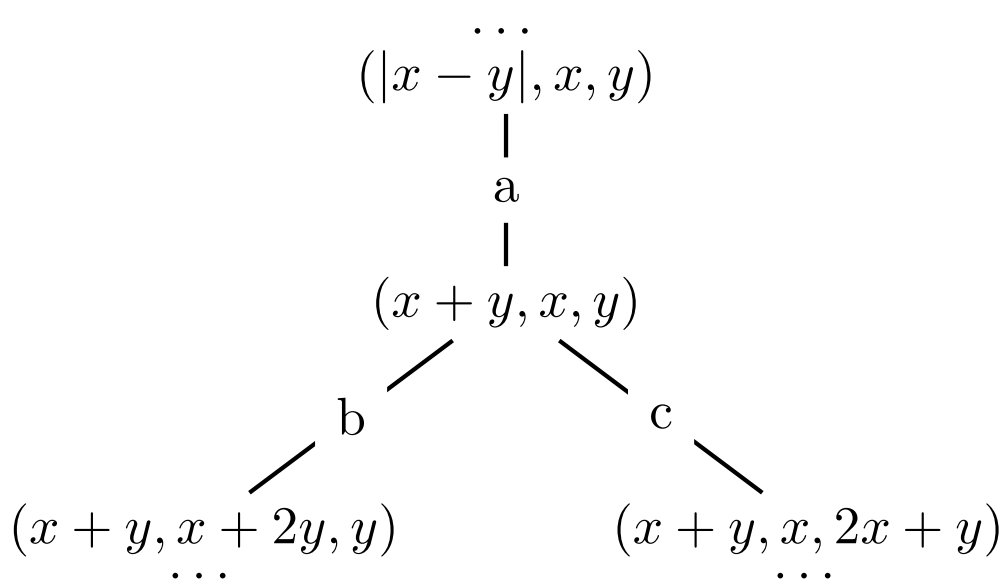
*Description $\mathcal{K}$ of epistemic model $\mathcal{T}$*:

1. $\bigvee_{x=y+z,y=x+z,z=x+y}(x_a \wedge y_b \wedge z_c)$
2. $\bigwedge_{y,z}((y_b \wedge z_c) \rightarrow K_a(y_b \wedge z_c))$
3. $\bigwedge_{y,z}((y_b \wedge z_c) \rightarrow (\hat{K}_a(y+z)_a \wedge \hat{K}_a(|y-z|)_a))$

*Description (characteristic formula) of an epistemic state $(\mathcal{T}, (x, y, z))$*:
$(x_a \wedge y_b \wedge z_c) \wedge C_{abc}\mathcal{K}$

# What is my number? — Structure of an $abc$-class

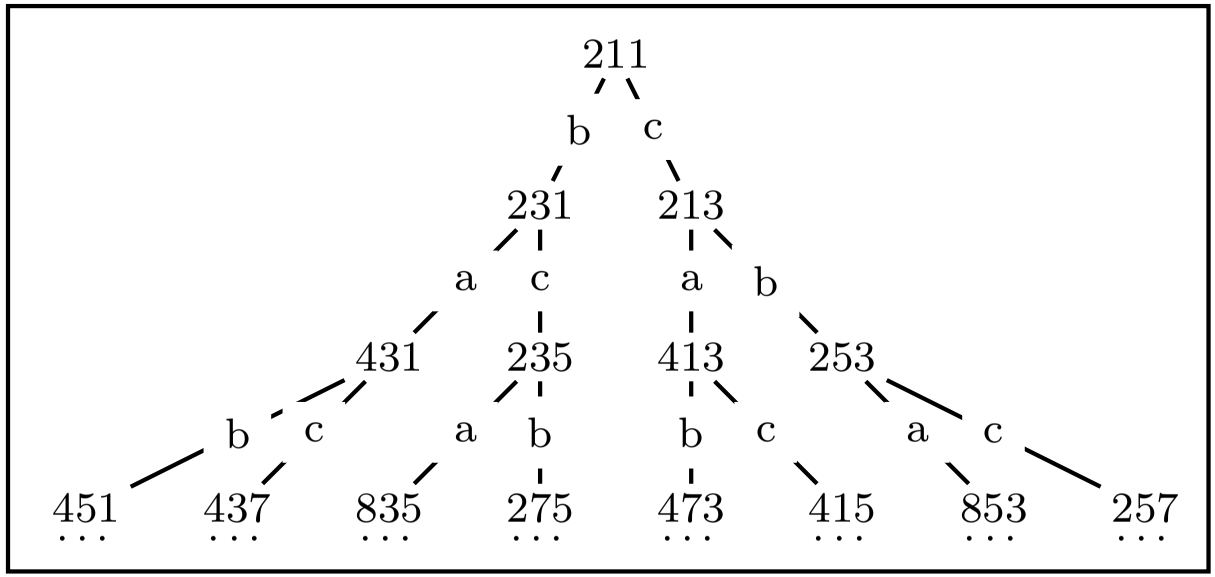
Background knowledge: what Anne, Bill, and Cath commonly know.
Branching in an arbitrary $abc$-equivalence class in the epistemic model $\mathcal{T}$.
(Arcs are labelled with the agent who cannot distinguish connected nodes.)



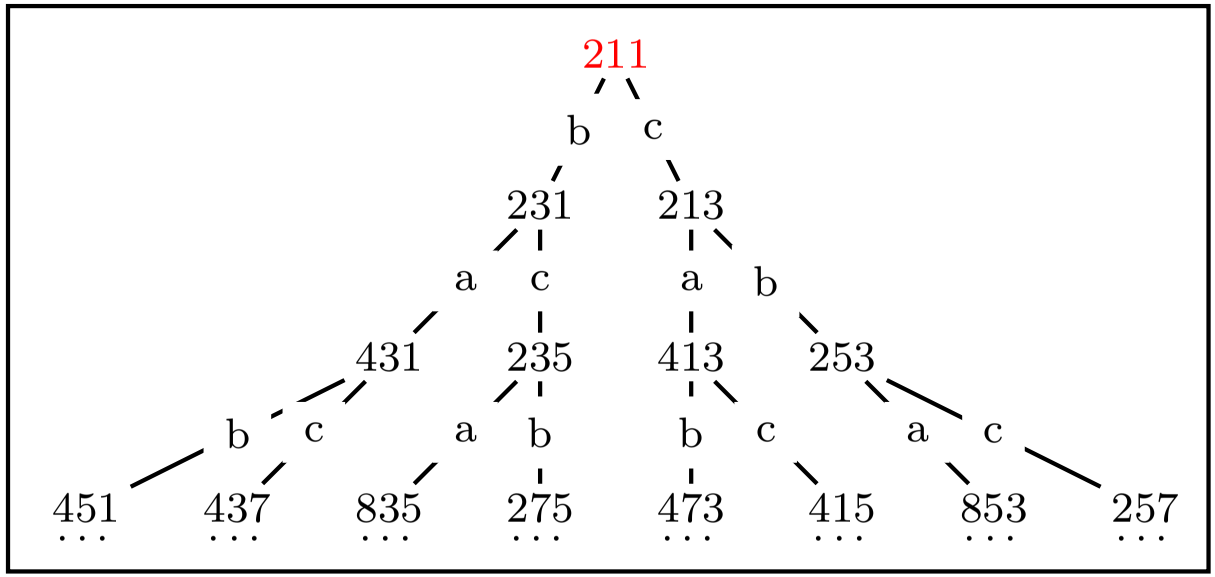
# What is my number? — Structure of an $abc$-class

All $abc$-classes look like $(2, 1, 1)$.

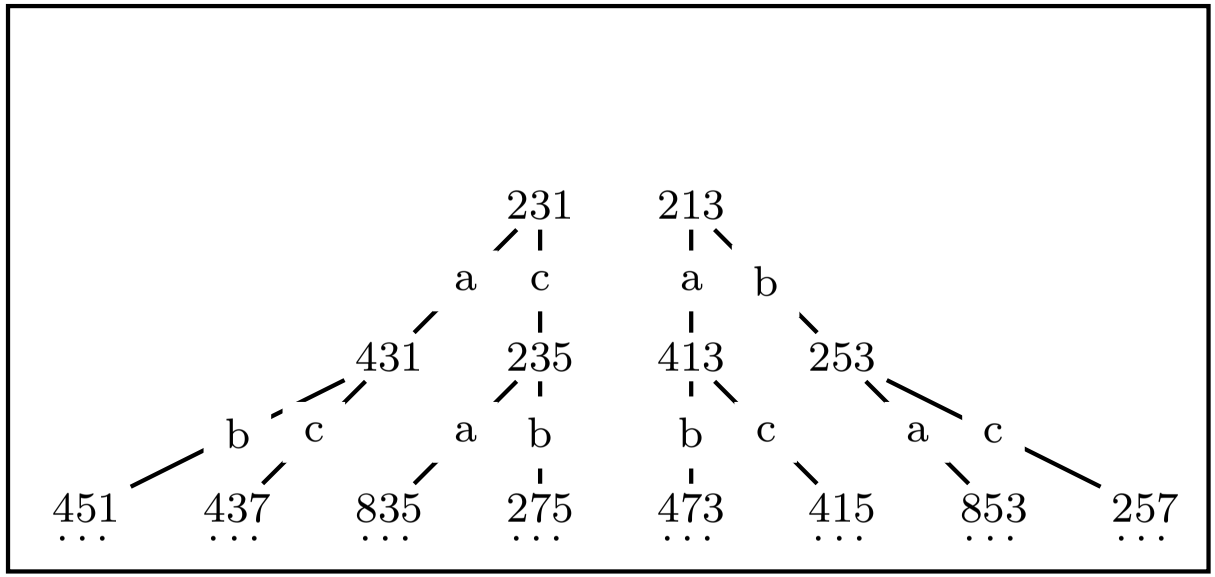
# Processing three ignorance announcements



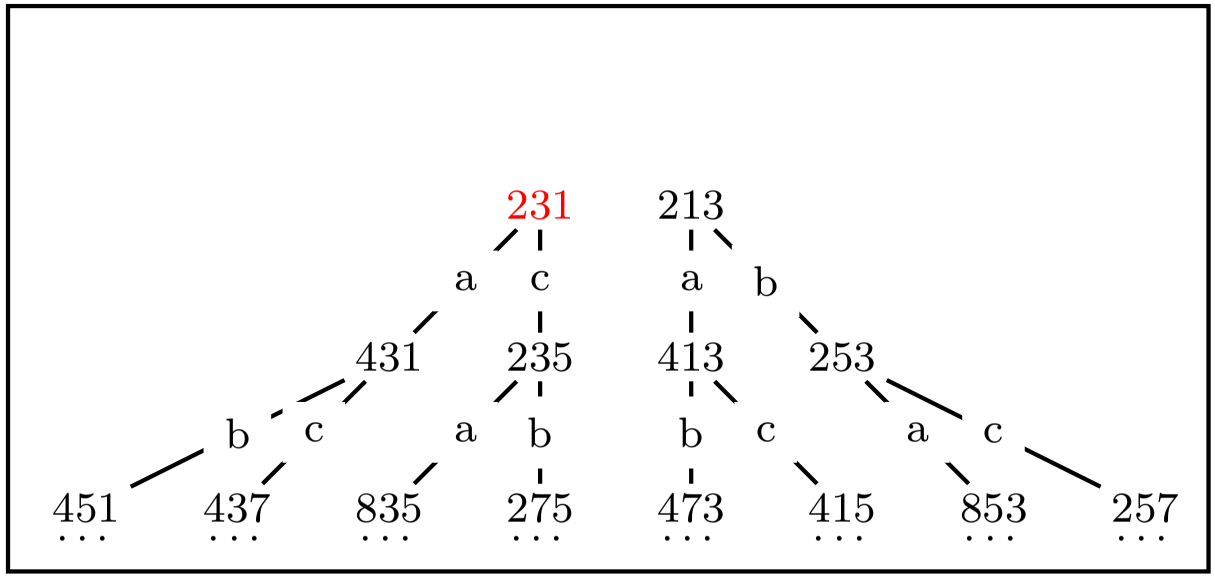
# Processing three ignorance announcements



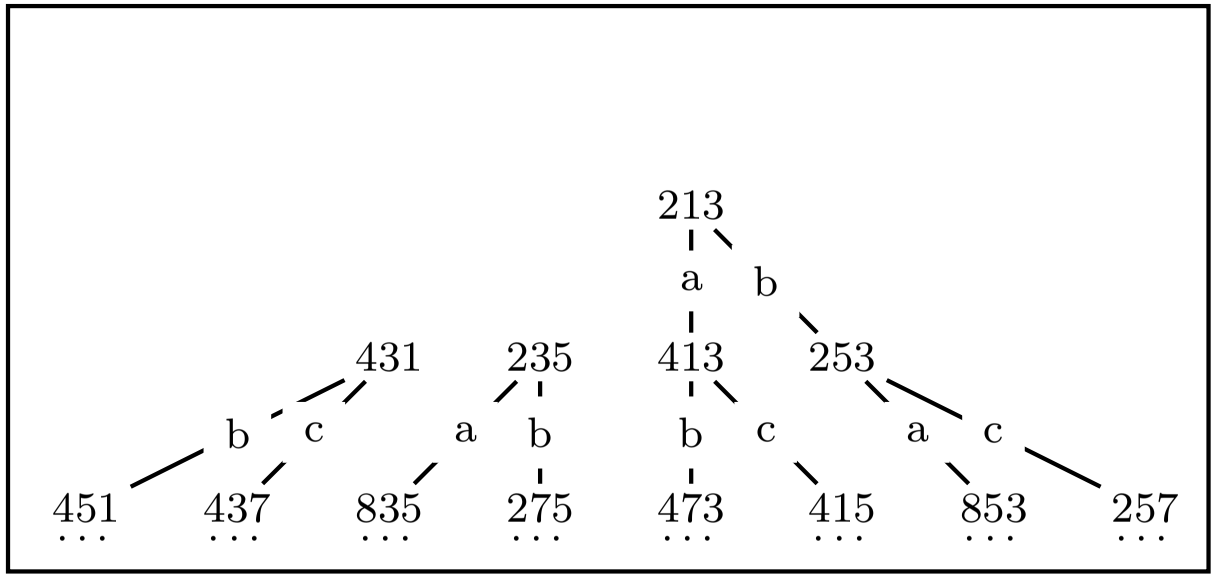
# Processing three ignorance announcements



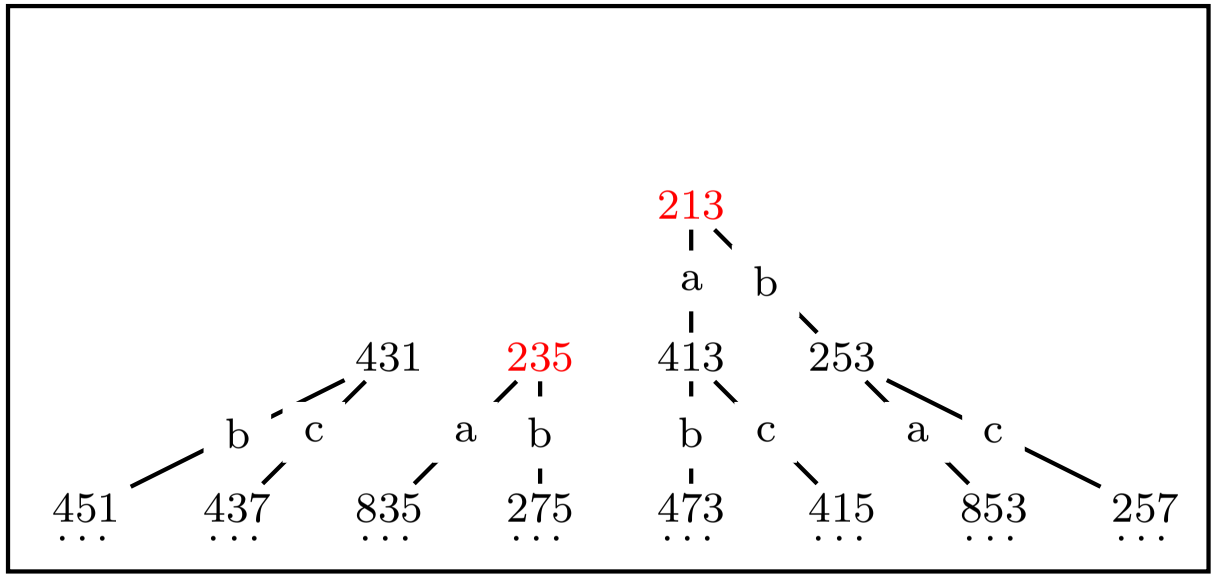
# Processing three ignorance announcements



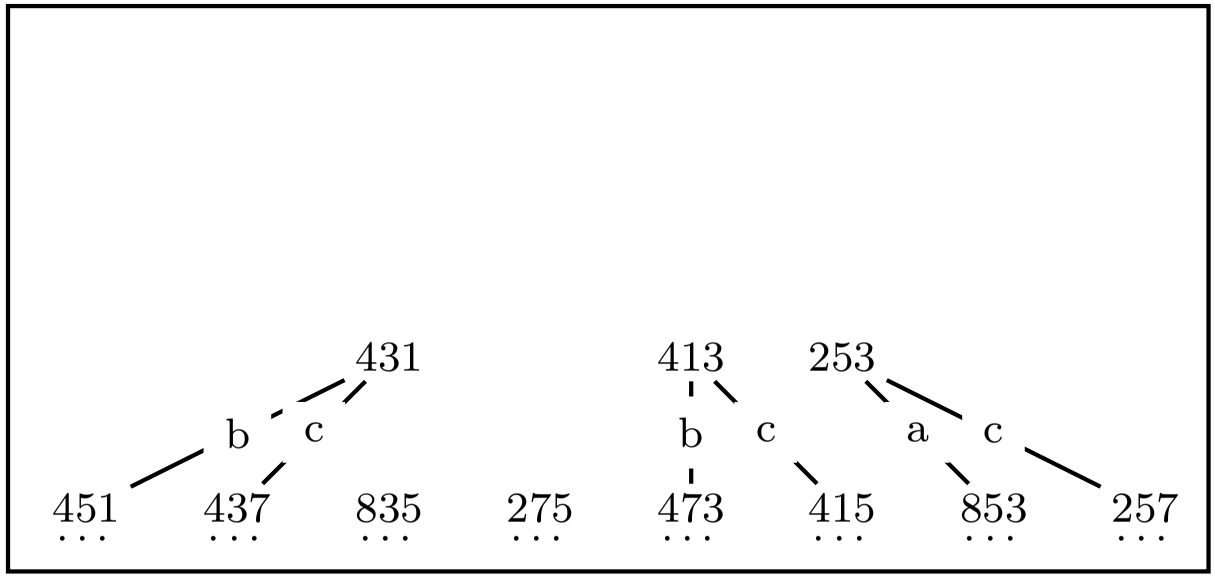
# Processing three ignorance announcements



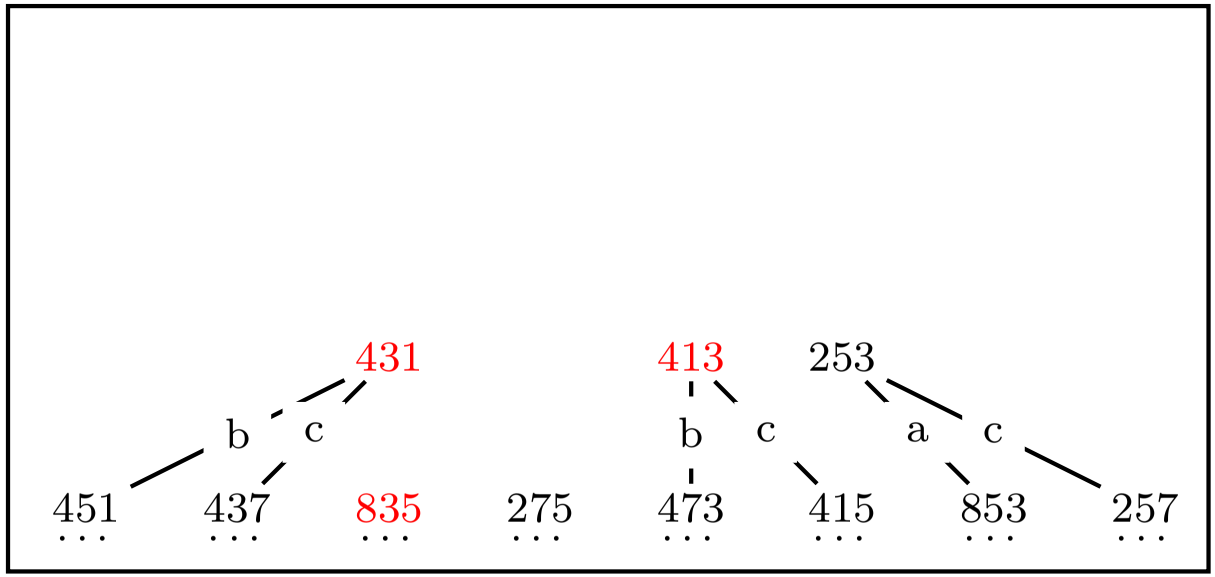
# Processing three ignorance announcements



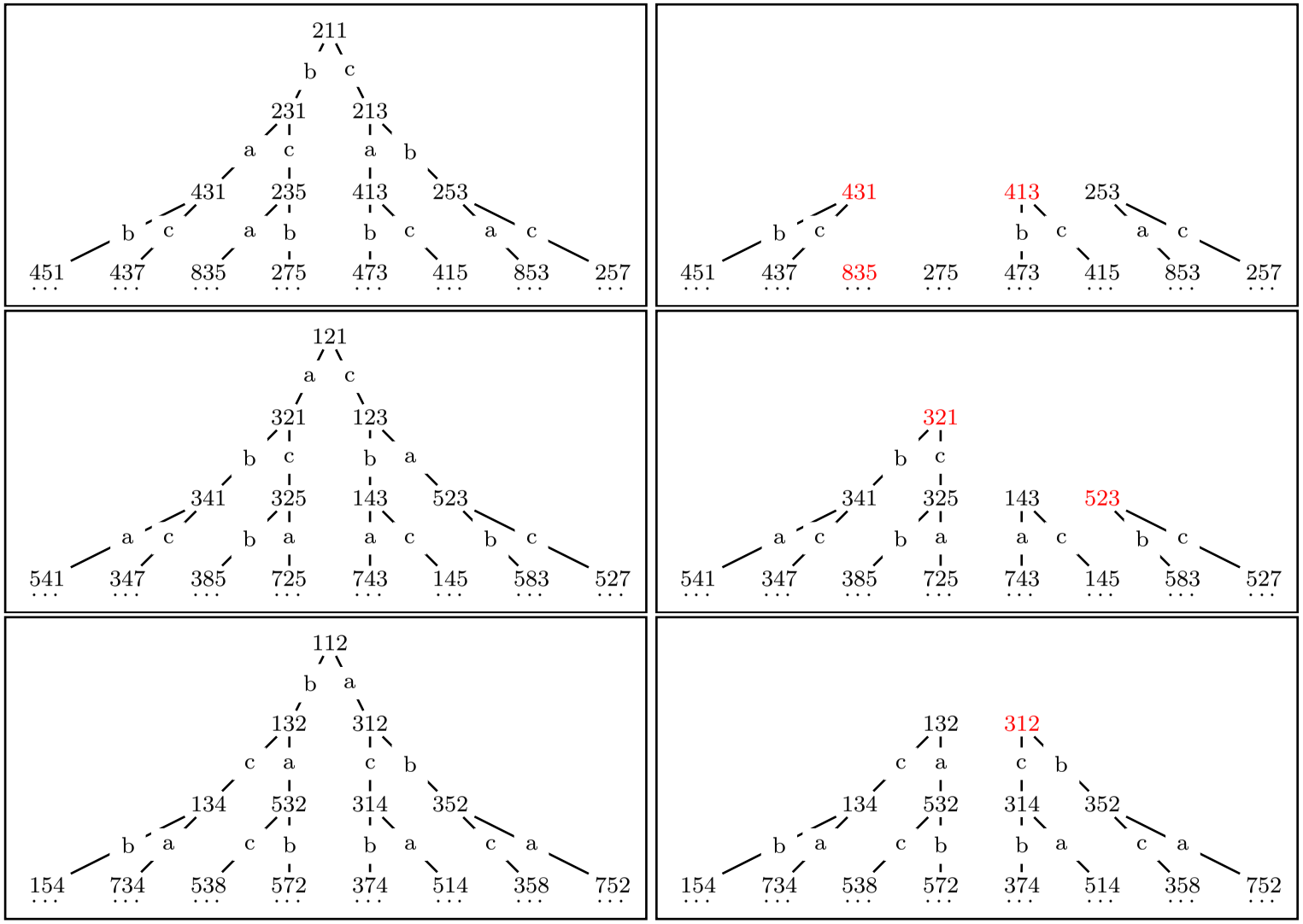
# Processing three ignorance announcements



# Processing three ignorance announcements



# Where after three ignorance announcements Anne knows

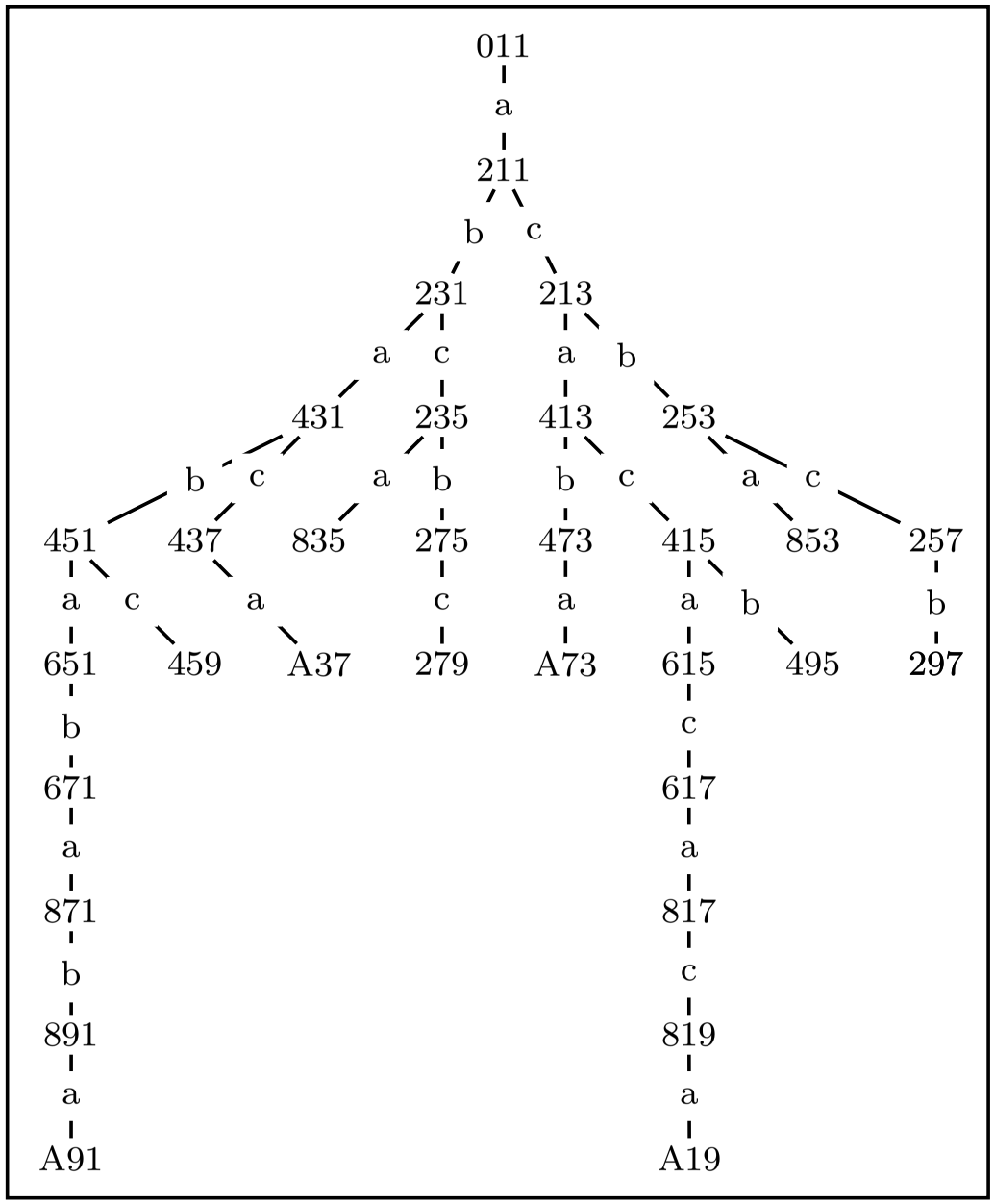


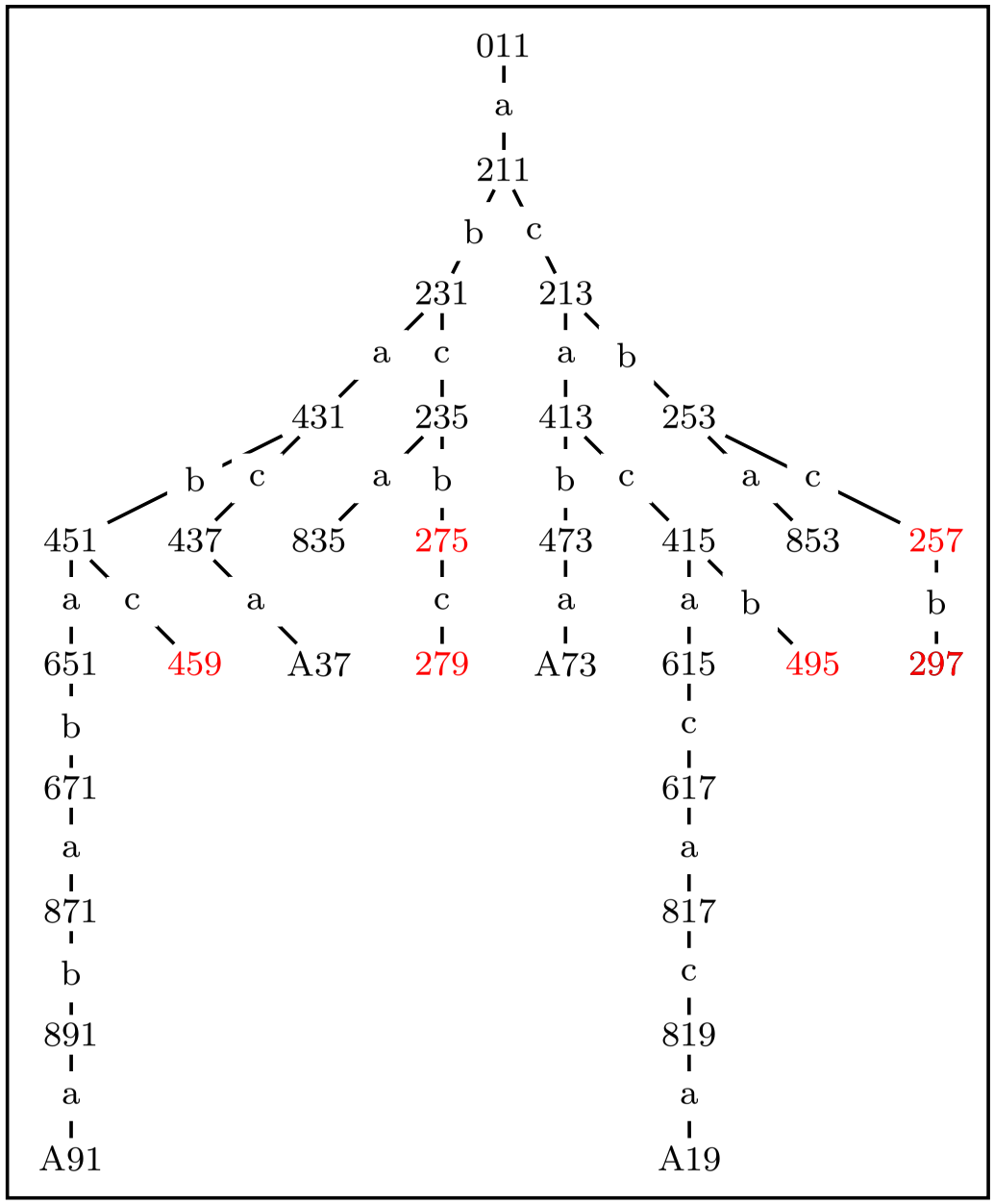
# What is my number? — different version of the riddle

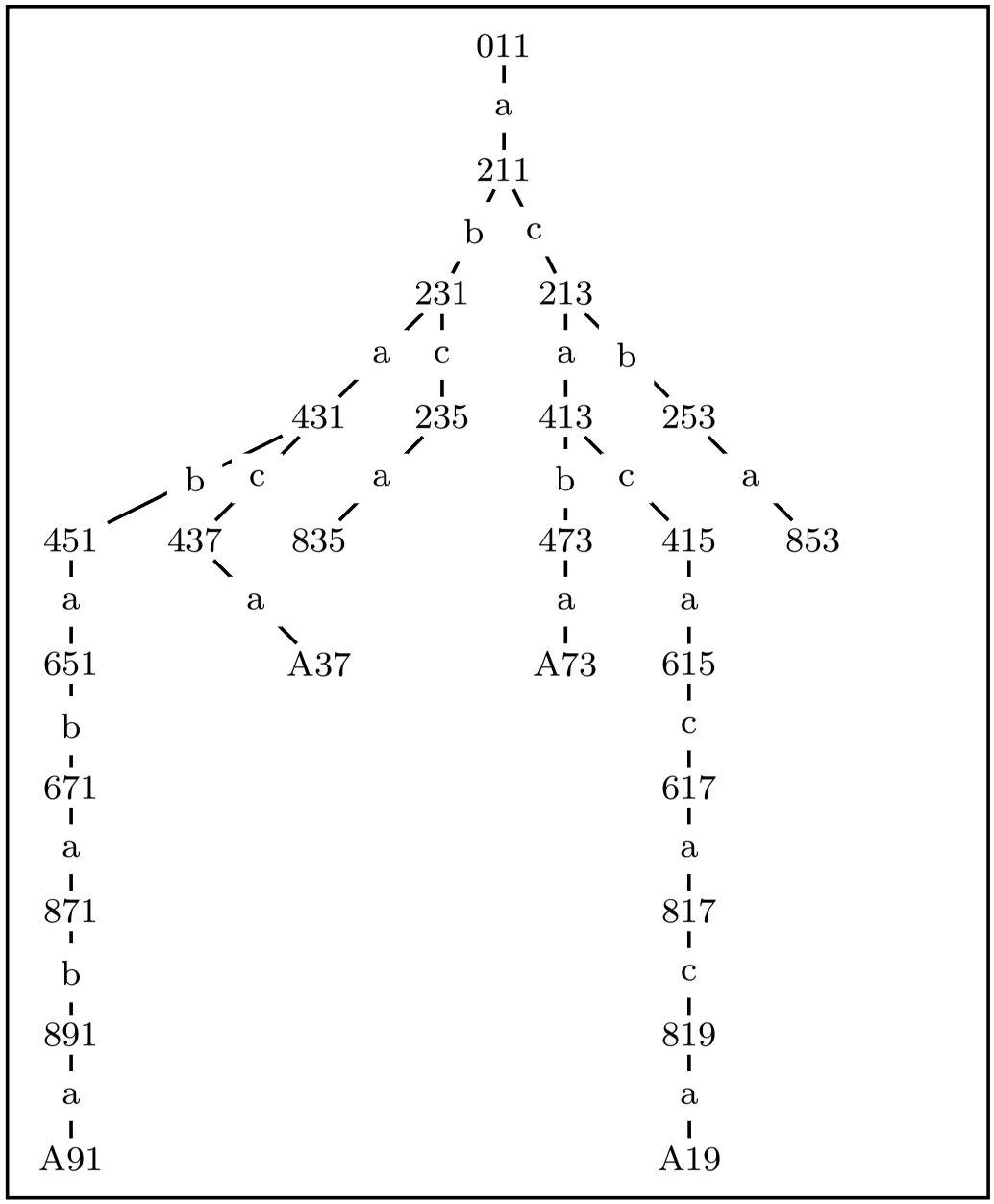
1. Anne: "I do not know my number."
2. Bill: "I do not know my number."
3. Cath: "I do not know my number."
4. Anne: "I know my number. ~~It is 50.~~"

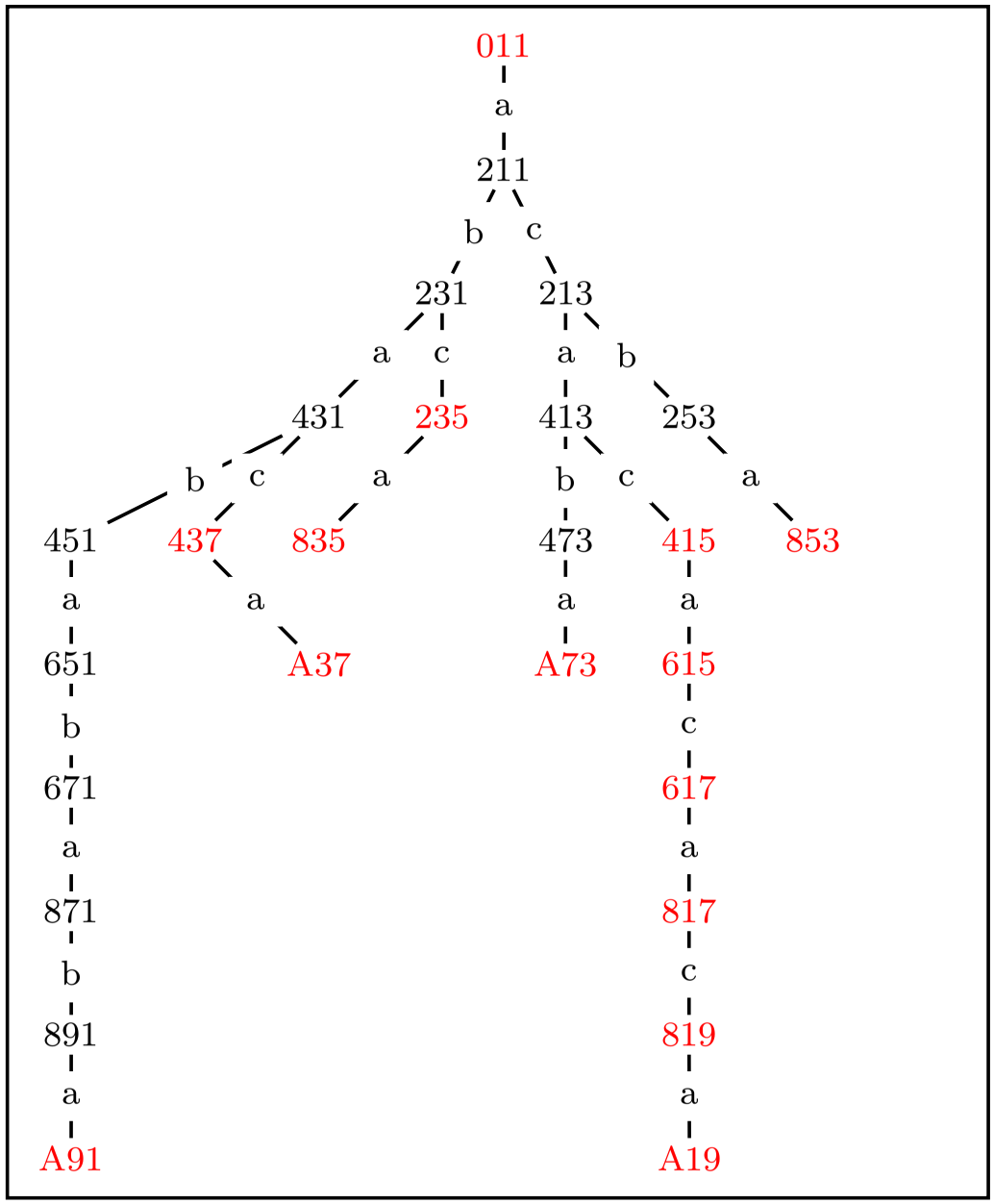
After 1, 2, 3, Anne *always* knows her number.
Not true for arbitrary numbers $x, y, z$.
Solvable for an upper bound **max** of $x, y, z$.

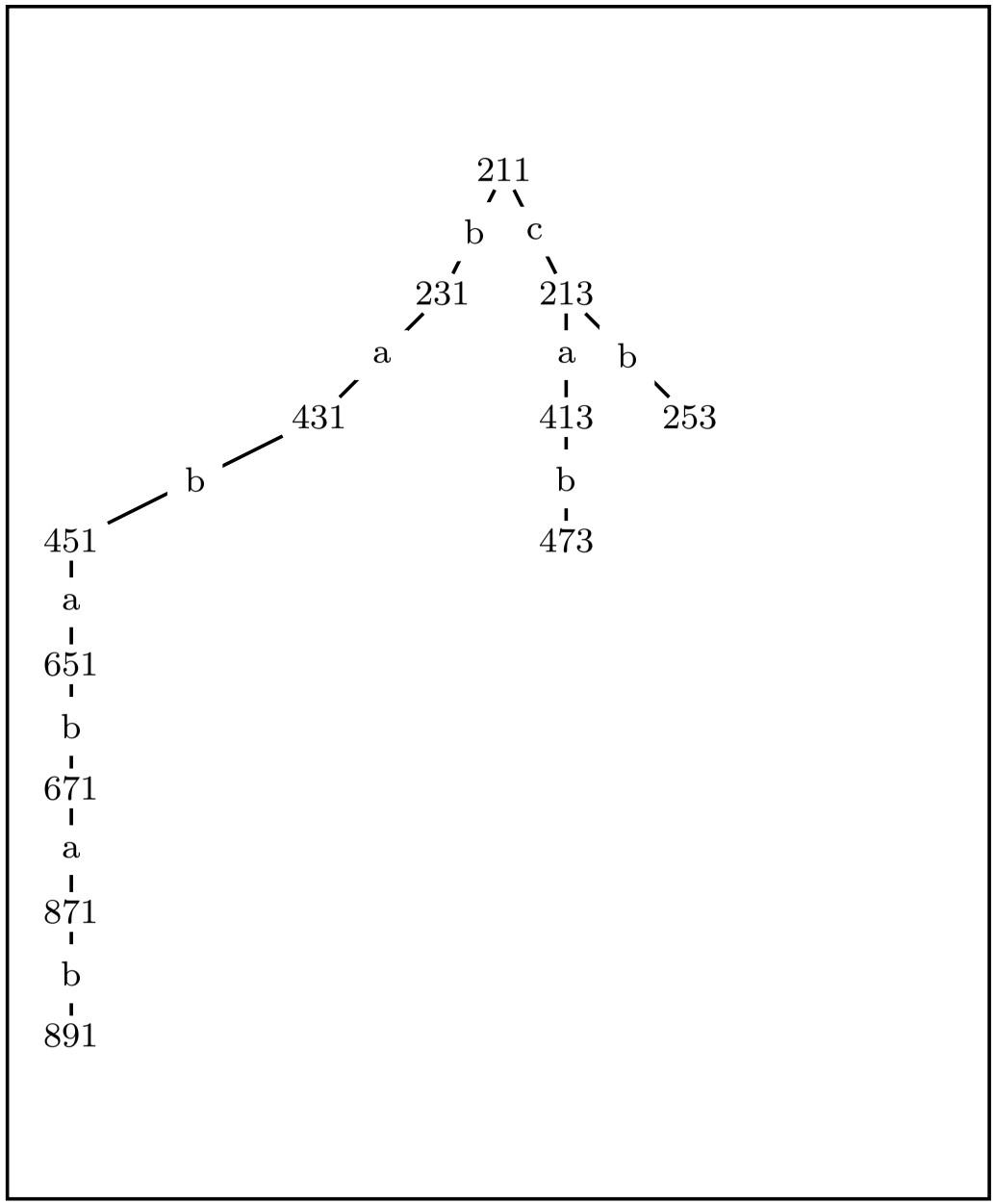
**What is the range of max, if Anne always knows her number?**
Suppose **max** $= 10$. What does the $abc$-class with triple $(0, 1, 1)$ look like?

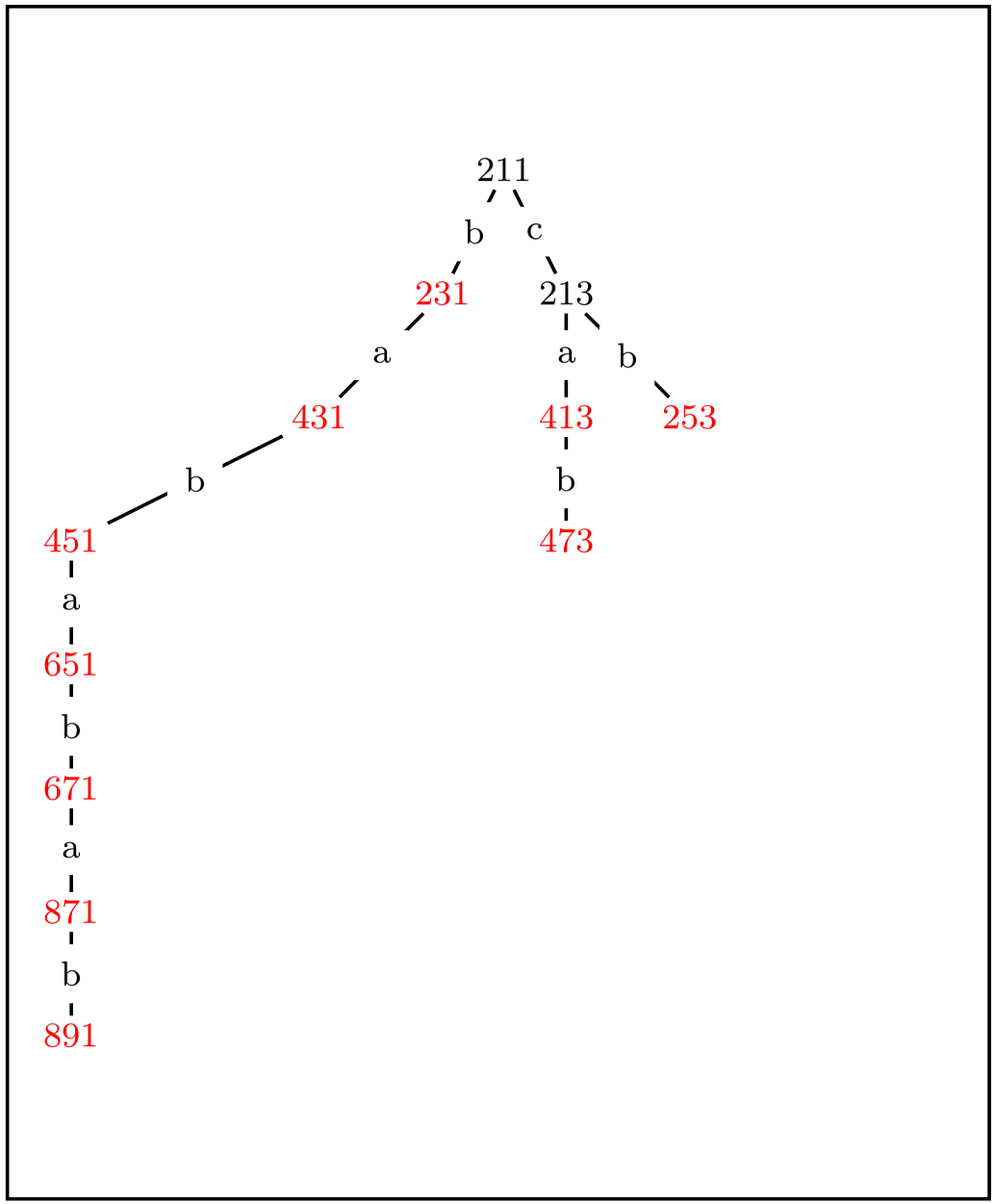

011
a
211
b c
231 213
a c a b
431 235 413 253
b c a b b c a c
451 437 835 275 473 415 853 257
a c a c a a b b
651 459 A37 279 A73 615 495 297
b c
671 617
a a
871 817
b c
891 819
a a
A91 A19

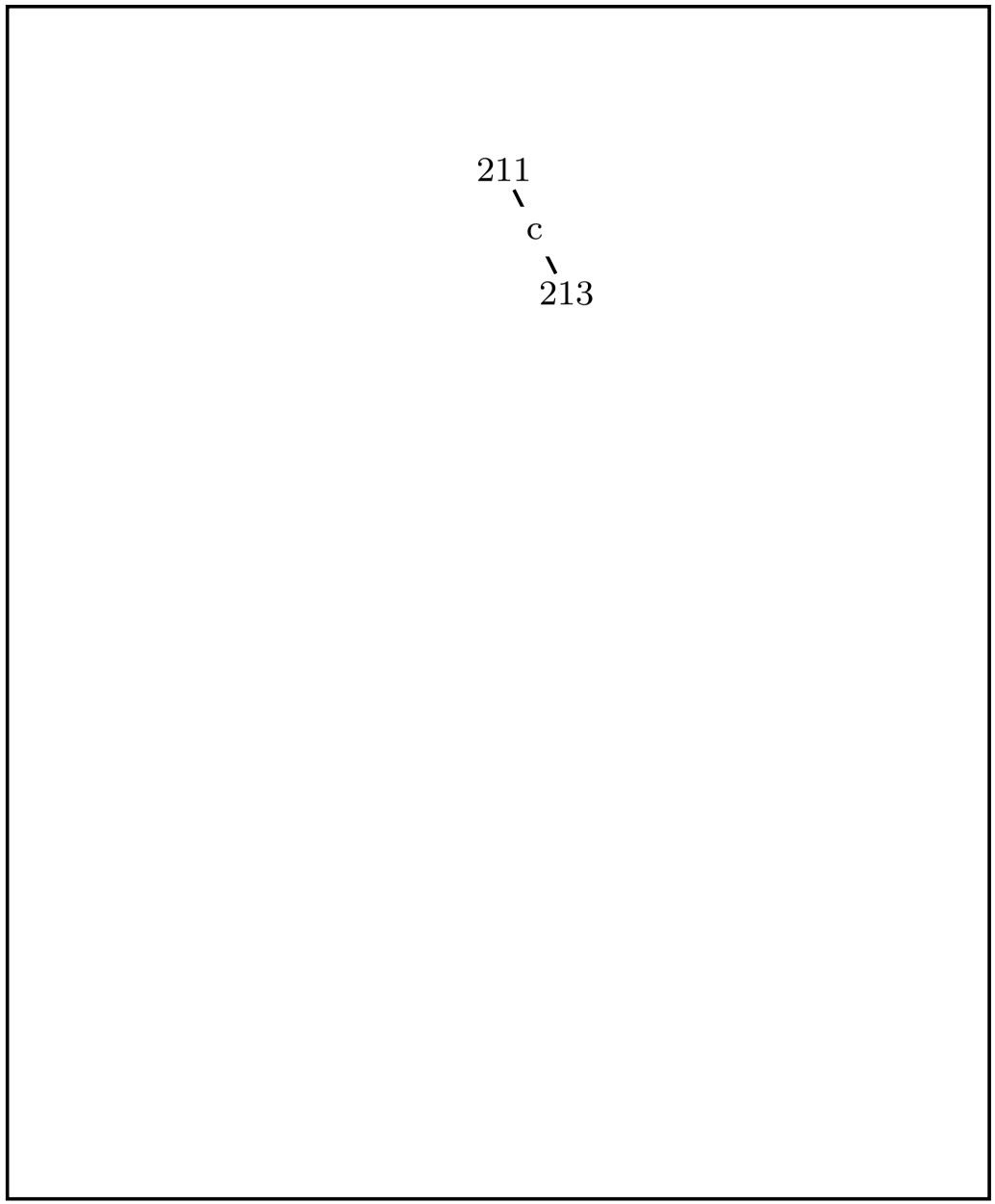

011
a
211
b c
231 213
a c a b
431 235 413 253
b c a b b c a c
451 437 835 275 473 415 853 257
a c a c a a b b
651 459 A37 279 A73 615 495 297
b c
671 617
a a
871 817
b c
891 819
a a
A91 A19


011
a
211
b c
231 213
a c a b
431 235 413 253
b c a b c a
451 437 835 473 415 853
a a a a
651 A37 A73 615
b c
671 617
a a
871 817
b c
891 819
a a
A91 A19


011
a
211
b c
231 213
a c a b
431 235 413 253
b c a b c a
451 437 835 473 415 853
a a a a
651 A37 A73 615
b c
671 617
a a
871 817
b c
891 819
a a
A91 A19


211
b
c
231
213
a
a
b
431
413
253
b
b
451
473
a
651
b
671
a
871
b
891


211
b
c
231
213
a
a
b
431
413
253
b
b
451
473
a
651
b
671
a
871
b
891


211
c
213

# What is my number? — different version of the riddle

Anne always knows her number, if $8 \leq \mathbf{max} \leq 13$.

*How about a verification tool?*

# Epistemic model checking

The epistemic model checker DEMO (in Haskell) implements action model logic.
**Action model logic**:
Baltag et al., *Logics for epistemic programs*, Synthese 139: 165–224, 2004.
**DEMO**:
J. van Eijck, *Dynamic Epistemic Modelling*, CWI Report SEN-E0424, 2004.

```
module SUMXYZ
where
import DEMO
upb = 10
-- triples (x,y,z) with x,y,z <= upb, x = y+z or y = x+z or z = x+y
triplesx  = [(x,y,z)|x<-[0..upb], y<-[0..upb], z<-[0..upb], x==y+z]
triplesy  = [(x,y,z)|x<-[0..upb], y<-[0..upb], z<-[0..upb], y==x+z]
triplesz  = [(x,y,z)|x<-[0..upb], y<-[0..upb], z<-[0..upb], z==x+y]
triples = triplesx ++ triplesy ++ triplesz
-- associating states with number triples
numtriples = llength(triples)
llength [] =0
llength (x:xs) = 1+ llength xs
itriples = zip [0..numtriples-1] triples
-- initial multi-pointed epistemic model
three :: EpistM
three = (Pmod [0..numtriples-1] val acc [0..numtriples-1])
 where
 val = [(w,[P x, Q y, R z]) | (w,(x,y,z))<- itriples]
 acc = [(a,w,v)| (w,(x1,y1,z1))<-itriples, (v,(x2,y2,z2))<-itriples, y1==y2, z1==z2]++
       [(b,w,v)| (w,(x1,y1,z1))<-itriples, (v,(x2,y2,z2))<-itriples, x1==x2, z1==z2]++
       [(c,w,v)| (w,(x1,y1,z1))<-itriples, (v,(x2,y2,z2))<-itriples, x1==x2, y1==y2]
-- agents a,b,c say (respectively): I do not know my number x,y,z
fagxnot = Conj [(Disj[Neg (Prop (P x)), Neg (K a (Prop (P x))) ])|x <-[0..upb]]
aagxnot = public (fagxnot)
fagynot = Conj [(Disj[Neg (Prop (Q y)), Neg (K b (Prop (Q y))) ])|y <-[0..upb]]
aagynot = public (fagynot)
fagznot = Conj [(Disj[Neg (Prop (R z)), Neg (K c (Prop (R z))) ])|z <-[0..upb]]
aagznot = public (fagznot)
-- model restriction resulting from the three announcements
solution = showM (upds three [aagxnot, aagynot, aagznot])
```

# Sum and Product

*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:*

*1. P says: "I do not know it."*

*2. S says: "I knew you didn't."*

*3. P says: "I now know it."*

*4. S says: "I now also know it."*

*Determine the pair $(x, y)$.*

**There is a unique solution!**

# Sum and Product — history

Originally stated, in Dutch, by Hans Freudenthal
*Nieuw Archief voor Wiskunde* 3(17):152, 1969.

Became popular in AI by way of John McCarthy.
*Formalization of Two Puzzles Involving Knowledge*, 1978-1981
In: Lifschitz, Formalizing Common Sense: Papers by John McCarthy, Ablex 1990.

Further contributions by Martin Gardner, I.M. Isaacs, ...

# Sum and Product

No. 223. $A$ zegt tot $S$ en $P$: Ik heb twee gehele getallen $x$, $y$ gekozen met $1 < x < y$ en $x + y \leqslant 100$. Straks deel ik $s = x + y$ aan $S$ alleen mee, en $p = xy$ aan $P$ alleen. Deze mededelingen blijven geheim. Maar jullie moeten je inspannen om het paar $(x, y)$ uit te rekenen.

Hij doet zoals aangekondigd. Nu volgt dit gesprek:

1. $P$ zegt: Ik weet het niet.
2. $S$ zegt: Dat wist ik al.
3. $P$ zegt: Nu weet ik het.
4. $S$ zegt: Nu weet ik het ook.

Bepaal het paar $(x, y)$.

(*H. Freudenthal*).

# Sum and Product — towards a solution

- if the number pair were $(2, 3)$:
  Product deduces the pair from their product.
- if the numbers were prime:
  Product deduces the pair: the unique factorization of the product.
- if the number pair were $(14, 16)$:
  their sum 30 is also the sum of 7 and 23; therefore, Sum cannot *know* that Product did not know the numbers.

Announcement 2 ($S$: "I knew you didn't.") supersedes the first.
The successive announcements 3 and 4 are also informative.

# Russian Cards
# Public communication of secrets

*From a pack of seven known cards* $0, 1, 2, 3, 4, 5, 6$ *Anne (a) and Bill (b) each draw three cards and Cath (c) gets the remaining card. How can Anne and Bill openly (publicly) inform each other about their cards, without Cath learning from any of their cards who holds it?*

Origin: Mathematics Olympiad, Moscow, 2000. (By way of Alexander Shen.)

# Public communication of secrets

> *From a pack of seven known cards* $0, 1, 2, 3, 4, 5, 6$ *Anne (a) and Bill (b) each draw three cards and Cath (c) gets the remaining card. How can Anne and Bill openly (publicly) inform each other about their cards, without Cath learning from any of their cards who holds it?*

Suppose Anne draws $\{0, 1, 2\}$, Bill draws $\{3, 4, 5\}$, and Cath 6.
Any announcement, or finite sequence of announcements, goes.

Suppose Anne says "I have 0 or 5." Anne cannot distinguish this from "I have 0 or 6". From that, Cath would learn that Anne has 0.

Suppose Anne says "I have 0 or 1 or 5." Cath can have at most one of these three cards. Therefore she remains uncertain about the ownership of the other two. But how to continue?

Suppose Anne says "I have 0 or 1 or two out of 2,3,4, and ..."
Can we explore such statements systematically?

# Public communication of secrets

Suppose Anne says "I have $\{0, 1, 2\}$, or Bill has $\{0, 1, 2\}$," and
Bill then says "I have $\{3, 4, 5\}$, or Anne has $\{3, 4, 5\}$".
After these two announcements, Cath appears not to know which of 012 or 345 is Anne's or Bill's hand. What is wrong with it?

Cath reasons "Suppose Anne does not have card 0. How can she know that her statement is true? She can't! Therefore she must have card 0. Same for 1 and 2. So Anne must hold 012."

# Structures for and logic of card deals

A deal of cards is a sequence of hands. Each hand is a set of cards.

It is known how many cards each player has.
This determines a domain $D$ of 'possible' card deals.
Players can only see their own hand of cards.
This induces an equivalence relation $\sim_a$ on the domain.

Epistemic state $(\langle D, \sim, V \rangle, d)$ represents card deal $d$.

# Structures for and logic of card deals

## Epistemic states for card deals

For the seven cards problem, we get $(\binom{7}{3} \cdot \binom{4}{3} =)$ 140 possible deals.

## Logical description

Fact $q_a$ describes that agent $a$ holds card $q$.
$ijk_a := i_a \wedge j_a \wedge k_a$ describes that $a$'s hand is $\{i, j, k\}$.

## Example

Suppose Anne draws $\{0, 1, 2\}$, Bill draws $\{3, 4, 5\}$, and Cath 6.
We write 012.345.6 for that card deal.
Anne's hand is described by $012_a$, etc.

# Public communication of secrets

The epistemic requirements appear to be that ($Q = \{0, 1, ..., 6\}$):

$$\begin{array}{ll} \mathsf{aknowsbs} & \bigwedge_{i \neq j \neq k \in Q} (ijk_b \rightarrow K_a ijk_b) \\ \mathsf{bknowsas} & \bigwedge_{i \neq j \neq k \in Q} (ijk_a \rightarrow K_b ijk_a) \\ \mathsf{cignorant} & \bigwedge_{q \in Q} \bigwedge_{a=a,b} \neg K_c q_a \end{array}$$

**Are these postconditions reached? Are they strong enough?**

Suppose Anne draws $\{0, 1, 2\}$, Bill draws $\{3, 4, 5\}$, and Cath 6.
Focus on Anne's announcement.

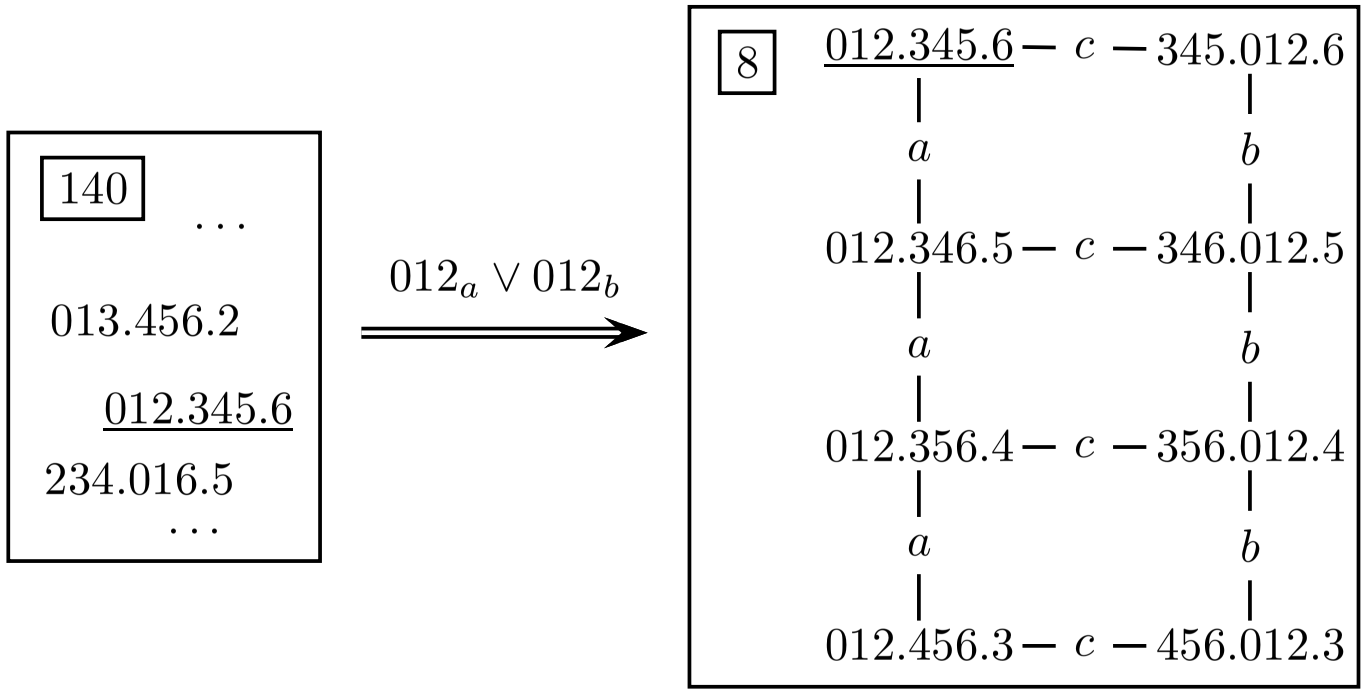
# Public communication of secrets: bad solution 1

*An insider says "Anne has* $\{0,1,2\}$ *or Bill has* $\{0,1,2\}$*."*
*Anne says "I have* $\{0,1,2\}$ *or Bill has* $\{0,1,2\}$*."*

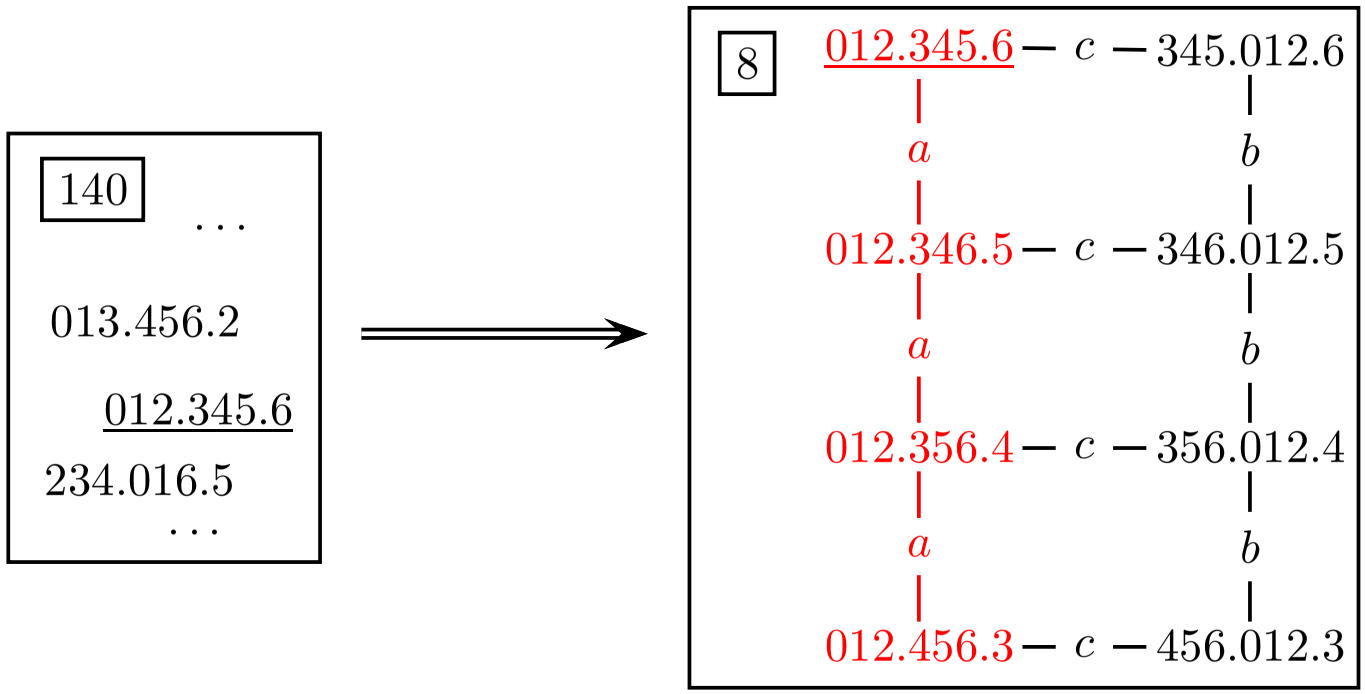
After the first, Cath remains ignorant.
After the second, she knows the entire deal of cards.
We may assume, that Anne *knows* what she says.

$$012.345.6 \models [012_a \vee 012_b]\mathsf{cignorant}$$
$$012.345.6 \not\models [K_a(012_a \vee 012_b)]\mathsf{cignorant}$$

# Public communication of secrets: bad solution 1



# Public communication of secrets: bad solution 1



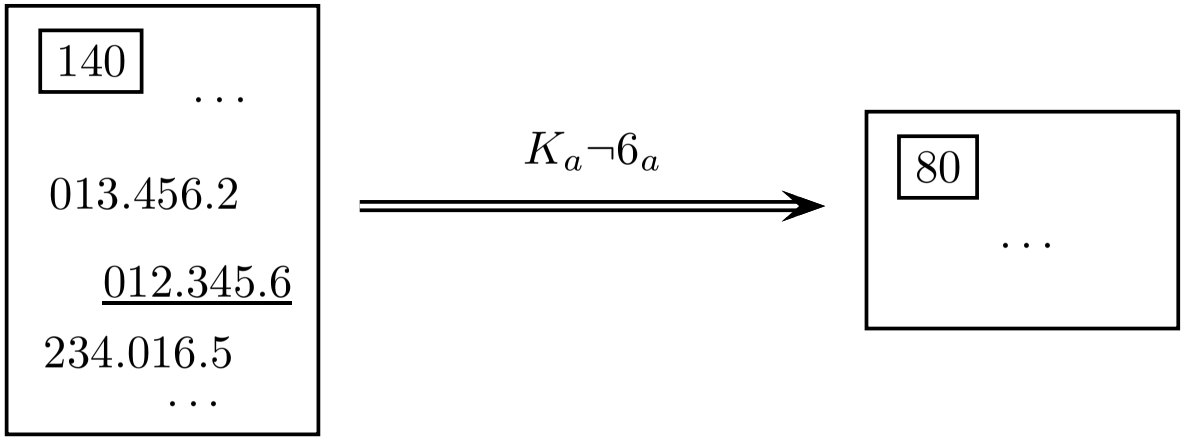
# Public communication of secrets: bad solution 2

*Anne says "I don't have card 6."*

In fact, Cath is ignorant after Anne's announcement.
Anne does not know whether Bill or Cath has card 6.
She will not risk that Cath learns a card of Bill.
We may assume that Anne *knows* that Cath is ignorant.

$$012.345.6 \models [K_a \neg 6_a]\mathsf{cignorant}$$
$$012.345.6 \not\models [K_a \neg 6_a] K_a \mathsf{cignorant}$$

# Public communication of secrets: bad solution 2



# Public communication of secrets: bad solution 3

*Anne says "I have* $\{0,1,2\}$*, or I have none of these cards."*

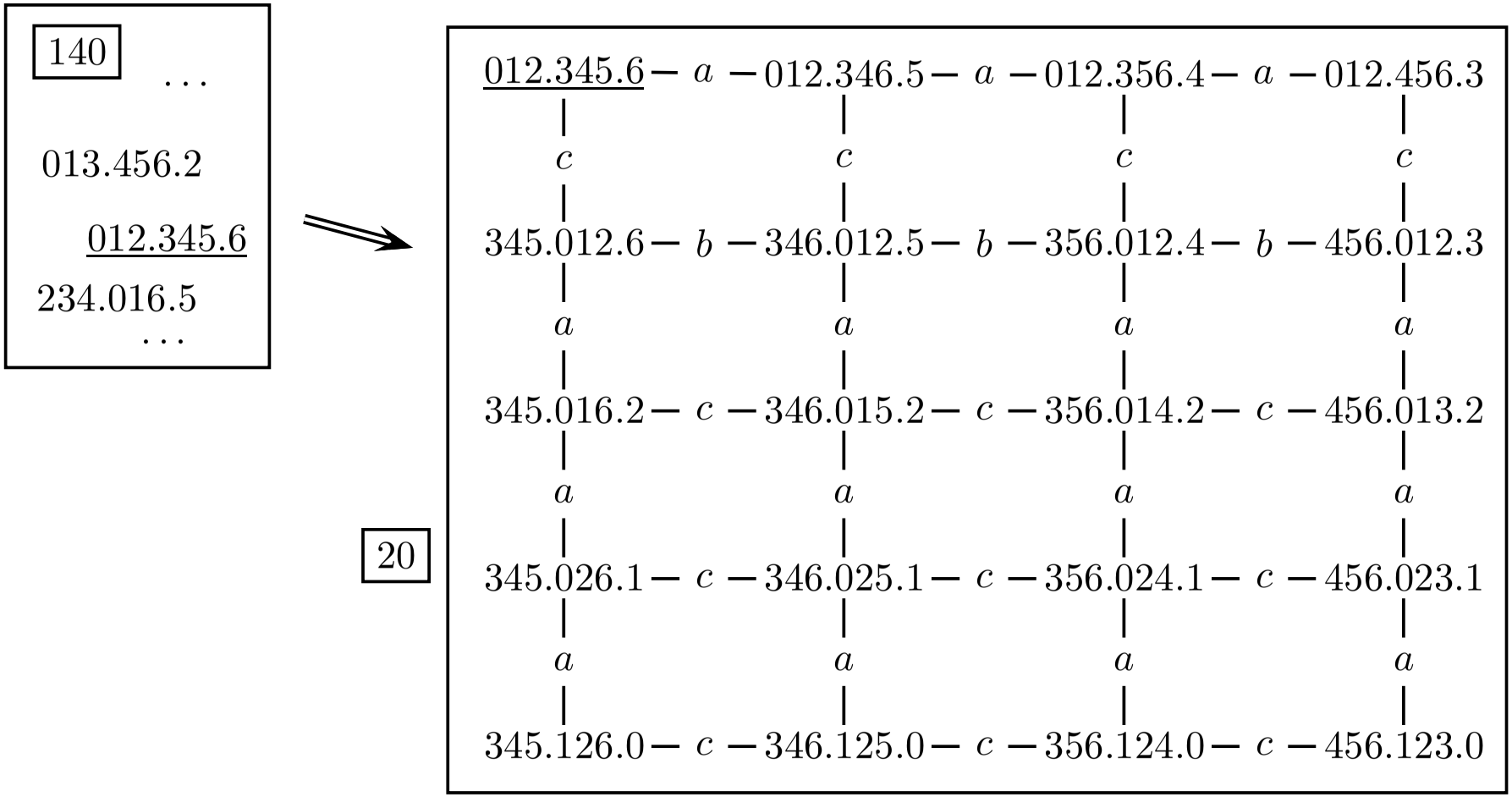
Cath is ignorant after Anne's announcement. Anne *knows* that Cath is ignorant.
But Cath doesn't know that Anne knows that she is ignorant.
We may assume that Anne knows that Cath is ignorant. But *that* is informative for Cath!

$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))]\mathsf{cignorant}$$
$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))]K_a\mathsf{cignorant}$$
$$012.345.6 \not\models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))]K_cK_a\mathsf{cignorant}$$
$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))][K_a\mathsf{cignorant}]\neg\mathsf{cignorant}$$

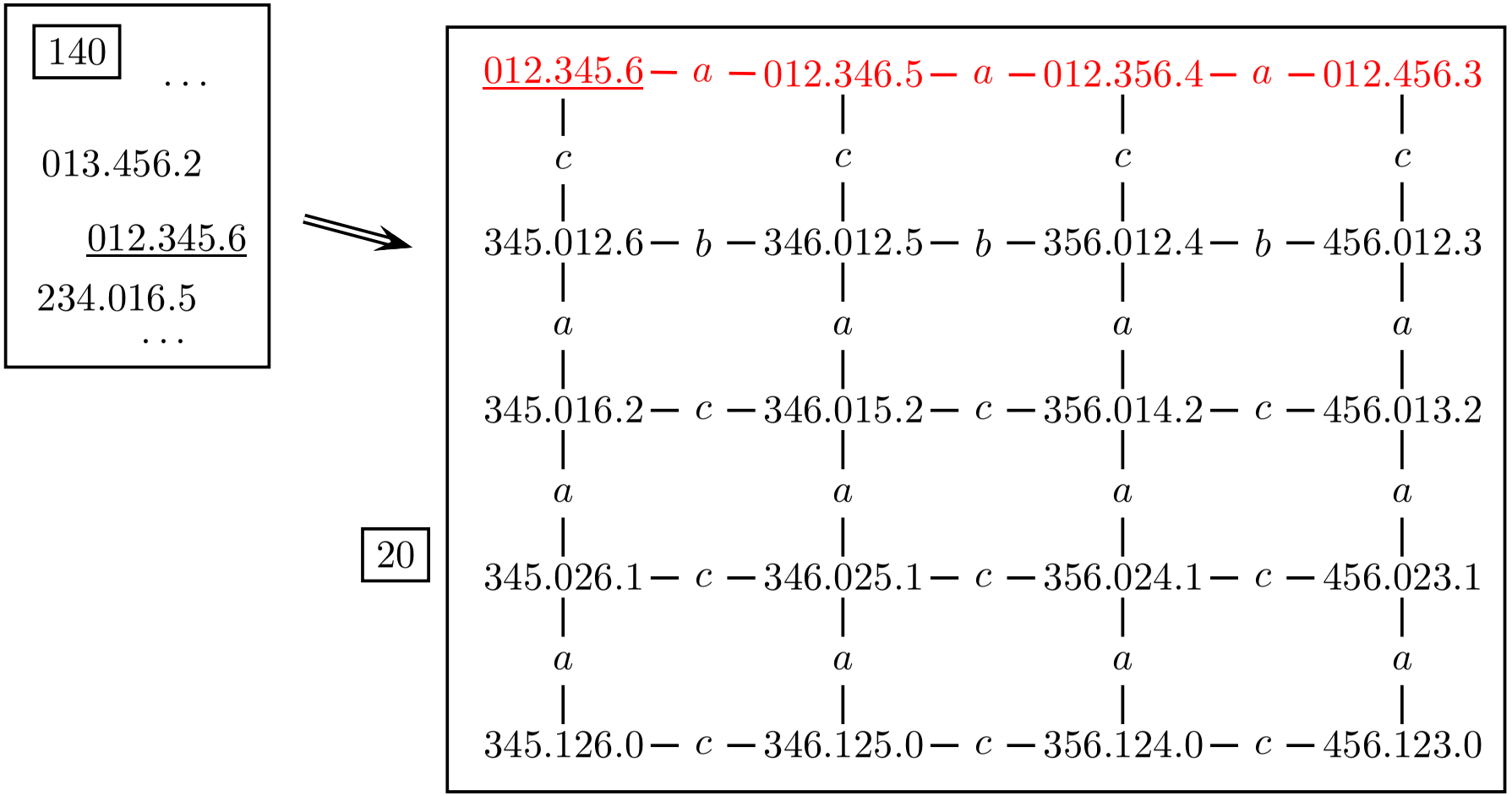
$$012.345.6 \models [K_a(012_a \vee \neg(0_a \vee 1_a \vee 2_a))][K_a\mathsf{cignorant}]\neg K_a\mathsf{cignorant}$$

$K_a\mathsf{cignorant}$ is an 'unsuccessful update' (formula becomes false when announced): Anne reveals her cards, *because* she intends to keep them secret.

# Public communication of secrets: bad solution 3



# Public communication of secrets: bad solution 3



# Public communication of secrets

*Safe announcements* guarantee absence of bad solutions:

| | |
|---|---|
| $[\varphi]$ | announcement of $\varphi$ (by an insider) |
| $[K_a\varphi]$ | announcement of $\varphi$ (by agent $a$) |
| $[K_a\varphi \wedge [K_a\varphi]C_A\mathsf{cignorant}]$ | safe announcement of $\varphi$ (by player $a$) |

An *exchange of secrets* is a sequence of safe announcements after which holds

$$C_{ab}(\mathsf{aknowsbs} \wedge \mathsf{bknowsas}) \wedge C_{abc}\mathsf{cignorant}$$

# Public communication of secrets: good solution 1

*Anne says "My hand is one of 012, 034, 056, 135, 246"* (anne).
*Bill says "Cath holds 6"* (bill).

After Anne's announcement it is common knowledge that Cath is ignorant.
After Anne's announcement it is common knowledge that Bill knows Anne's hand.
After Bill announces Cath's card, all three requirements are commonly known.

$$012.345.6 \models [\mathsf{anne}](\mathsf{bknowsas} \wedge C_{abc}\mathsf{cignorant})$$
$$012.345.6 \models [\mathsf{anne}][\mathsf{bill}]C_{abc}(\mathsf{aknowsbs} \wedge \mathsf{bknowsas} \wedge \mathsf{cignorant})$$

# Public communication of secrets: good solution 2

*Anne says "My hand is one of 012, 034, 056, 135,* **245**, *246."*

We now achieve $C_{ab}(\mathsf{aknowsbs} \wedge \mathsf{bknowsas})$ but not $C_{abc}(\mathsf{aknowsbs} \wedge \mathsf{bknowsas})$. Cath considers it possible that the deal is 245.013.6. In that case, Bill would not have learnt Anne's hand. Anne considers that possible.

$$
\begin{aligned}
012.345.6 &\models [\mathsf{anne}] C_{abc} \mathsf{cignorant} \\
012.345.6 &\models [\mathsf{anne}] \mathsf{bknowsas} \\
012.345.6 &\models [\mathsf{anne}] K_a \mathsf{bknowsas} \\
012.345.6 &\not\models [\mathsf{anne}] K_c K_a \mathsf{bknowsas} \\
245.013.6 &\not\models [\mathsf{anne}] K_a \mathsf{bknowsas} \\
245.013.6 &\models [\mathsf{anne}] \neg \mathsf{bknowsas}
\end{aligned}
$$

Even if Cath eavesdropped on Anne and Bill, she cannot be sure Bill now has the secret. Will she break into Bill's computer, or not, to get the secret?

For the solution, see JANCL 'The case of the hidden hand', 2005.

# Cryptology for ideal agents: theoretical issues

- Design closed systems where the probability of guessing the secret correctly is below some threshold (5%, 1 %, .5 %, ...)
- Are there protocols of length strictly larger than 2? (I.e., more than one message from sender and one from receiver.) Work by Fischer/Wright, Nishizeki et al. suggests longer protocols (for arbitrary bit exchange).
- Does the length of the protocol provide bias that the eavesdropper can benefit from?
- Can bias in card occurrence be used by the eavesdropper?
- Does this generalize to other (closed!) interpreted systems?
- Protocols where not just information but also *facts* (such as keys) change.

# Model checking

- MCK - Gammie and van der Meyden
- MCMAS - Lomuscio and Raimondi
- DEMO - van Eijck
- Very simple protocols create huge computational difficulties: the 'good solution 1' consisting of five hands 012 034 056 135 246 was implemented and verified in all three. Beyond that, trouble on the horizon.

See 'Model Checking Russian Cards' (v Ditmarsch, vd Hoek, vd Meyden, Ruan), ENTCS 2006.

# Cryptology for ideal agents: applications

- Any closed system where scarce known resources are distributed over agents. Instead of cards, it can be print jobs, locations (in which of seven locations in Afghanistan is Bin-Laden), ...
- Does this generalize to open systems?
- Redescription of known protocols in information-based terms and model checking their properties.

# Russian Cards / References – others


- M.J. Fischer and R.N. Wright. *Bounds on secret key exchange using a random deal of cards.* Journal of Cryptology, 9(2):71–99, 1996.
- T. Mizuki, H. Shizuya, and T. Nishizeki. *A complete characterization of a family of key exchange protocols.* International Journal of Information Security, 1:131–142, 2002.
- R. Ramanujam and S.P. Suresh. *Information based reasoning about security protocols.* Electronic Notes in Theoretical Computer Science, 55(1), Elsevier, 2003.
- A. Stiglic. *Computations with a deck of cards.* Theoretical Computer Science, 259: 671–678, 2001.

# Russian Cards / References – own


- H.P. van Ditmarsch. *The Russian Cards Problem.* Studia Logica 75: 31-62, 2003.
- M.H. Albert, R.E.L. Aldred, M.D. Atkinson, H.P. van Ditmarsch, C.C. Handley. *Safe communication for card players by combinatorial designs for two-step protocols.* Australasian Journal of Combinatorics, 33:33–46, 2005.
- H.P. van Ditmarsch, W. van der Hoek, R. van der Meyden and J. Ruan. *Model Checking Russian Cards.* Electronic Notes in Theoretical Computer Science 149:105–123, 2006.
- H.P. van Ditmarsch. *The case of the hidden hand.* Journal of Applied Non-Classical Logics 15/4:437–452, 2005.

# ‘Darkness at Noon’

# A bit of Arthur Koestler, with a lot of Moshe Vardi

*Russia in its darkest hour... A group of prisoners, meeting all together in the prison dining area, are told that: After dinner they will all be put in isolation cells. They will then be interrogated one by one in a room containing a single table-lamp with an on/off switch. The lamp is currently switched off (and only prisoners can manipulate the light-switch), there is no fixed order of interrogation and the same prisoner may be interrogated again at any stage. If the prisoners can find out that they all have been interrogated at least once, they will all be set free, immediately, but in case they mistakenly claim to know that, they will all be killed, also immediately.*

*Can the prisoners find out whether they have all been interrogated?*