The power of saying the obvious

Lie Fu

University of Strasbourg

25 March, 2026, EES

A tragedy on an island

An isolated island

- There is an island with a group of very smart inhabitants.
- Some of them have red eyes ; the others do not.
- Every islander can see everyone else's eyes, but **cannot see their own**.
- There are no mirrors, no cameras, and no reflective surfaces.
- All islanders are perfectly logical, and they are aware of this.

Important

Everyone sees the world in the same way — except for one missing piece :

"What color are my own eyes ?"

A religion – the rules

- On the island, it is forbidden to talk about eye colors.
- e.g. they never tell each other their eye color.
- If an islander ever becomes certain of their own eye color, he/she must exile himself/herself and leave the island that night.
- Everyone will know if someone leaves the island.

A visitor arrived at the island...

... and said to everyone in a gathering :

Visitor (public announcement) :

"Some of you have red eyes !"

Everyone thinks : Oh, no ! We don't talk about this subject here. But let's forgive him. After all, what he said is **obvious** ! Everyone knows...

- Nothing happened in the first day.
- Nothing happened in the second day.
- ...
- Nothing happened in the first few days

(Why would anything happen ? The visitor's annoucement is not news.)

But...

The tragedy

After 10 days, a group of islanders exiled themselves and left at the same night !

Question for you – my detectives :

What happened ? Why ?

- What the visitor said must changed something.
- Why should this announcement matter ? He only said the obvious.

George Pólya

If you can't solve a problem, then there is an easier problem you can solve : find it.

– Mathematical Discovery on Understanding, Learning, and Teaching Problem Solving

The easiest cases

- If there is **1** red-eyed islander, what happens ?
- If there are **2** red-eyed islanders, what happens ?
- If there are **3** red-eyed islanders, what happens ?

2 red-eyed case

- Alice sees Bob with red eyes
- Bob sees Alice with red eyes
- Alice thinks : If I do not have red eyes, Bob would be the only one, and Bob would leave on night 1
- Bob reasons the same way
- night 1 passes and nobody leaves
- so each concludes : I must also have red eyes
- both leave on night 2

They both implicitly used : *Alice knows that Bob knows* that there are people some red-eyes.

3 red-eyed case

- each sees two red-eyed people
- each thinks : "If I don't have red eyes, then those two are in the 2-person case, so they should leave on night 2"
- when they do not, each concludes they must also be red-eyed
- all three leave on night 3

They all implicitly used : *A knows that B knows that C knows* that there are people some red-eyes.

From patterns to conjectures and theorems

The conceptual "Giant Leap" : $1 \to 2 \to 3 \longrightarrow$ all n

Conjecture

If there are n red-eyed islanders, they all leave on night n.

This is a mathematical statement, depending on a natural number – an infinity set.

- examples suggest the pattern/conjecture, but not enough to *prove* it.
- Method of proof : **Induction**.

Induction

Induction is a method to prove a statement of the form :

"For any natural number n, $P(n)$ is true."

Procedure : two steps :

(1) (Initial case) : check that $P(1)$ is true.

(2) (Induction) : prove that $P(n)$ is true under the *assumption* that $P(n-1)$ is true. In short : $P(n-1) \Longrightarrow P(n)$

Then we can conclude $P(n)$ is true for any n.

An example of induction

Prove that for any $n \geq 1$,

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}.$$

(1) (Initial case) : $1 = \frac{1\times(1+1)}{2}$

(2) (Induction) : Assuming that $1 + 2 + \cdots + (n-1) = \frac{n(n-1)}{2}$, then $1 + 2 + \cdots + n = (1 + 2 + \cdots + (n-1)) + n = \frac{n(n-1)}{2} + n = \frac{n(n+1)}{2}$.

Exercise : $1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}$.

Back to the islander puzzle

Prove by induction the following statement $P(n)$:

"If there are n red-eyed islanders, then they all leave at night n."

(1) (Initial case $P(1)$) : If there is only 1 red-eyed, he/she realizes his eye color immediately after visitor's speech, and leaves the island the first night.

(2) (Induction $P(n-1) \Longrightarrow P(n)$) : Assuming : "If there are $(n-1)$ red-eyed islanders, then they all leave at night $n-1$."
Now if there are n red-eyed islanders. Each of them will think :"If I don't have red eyes, then there are $n-1$ red-eyed islanders, so **using the assumption** $P(n-1)$, all of them should leave on night $n-1$".
But they all think like this and no one left on night $(n-1)$.
When they observed this, each red-eyed islander deduced their eyes' color, and left on night n. We proved $P(n)$.

The real mystery

We now understand *what* happens.

But why ?

Why does the visitor's announcement matter at all ?

After all, if there are at least two red-eyed islanders, then everyone can already *see* at least one of them and the visitor's annoucement is not new information.

Key question

What changes when a fact is said aloud in **public** ?

Different levels of knowing

Let P be the knowledge

$$P = \text{“At least one islander has red eyes.”}$$

There are several different ways in which P can be known.

- P is true.
- Every islander knows P.
- Every islander knows that every islander knows P.
- Every islander knows that every islander knows that every islander knows P.
- ...

Important

These are *not* the same thing. In the proof by induction, we used **all** of the levels.

Mutual knowledge

Again let P be the knowledge/statement

"At least one islander has red eyes."

Mutual knowledge

P is mutual knowledge if every islander knows P.

In the red-eyed islanders puzzle, before the visitor speaks :

- if you see someone with red eyes, then you know P ;
- in fact, every islander knows P.

But

Mutual knowledge only says :

"Everyone knows P."

It says *nothing* about what people know about other people's knowledge.

Common knowledge

The idea of common knowledge is much stronger.

Common knowledge

A statement P is common knowledge if :

- everyone knows P,
- everyone knows that everyone knows P,
- everyone knows that everyone knows that everyone knows P,
- and so on, without end.

Common knowledge = knowledge at *all levels*

Mutual knowledge vs. common knowledge

Mutual knowledge

- Everyone knows P
- One level only
- No information about

  "who knows that others know"

Common knowledge

- Everyone knows P
- Everyone knows that everyone knows P
- Everyone knows that everyone knows that everyone knows P
- And so on forever

Mutual knowledge is *much weaker* than common knowledge.

Why the announcement changes everything

Before the visitor speaks, the fact P may already be *mutual knowledge*.

But after the visitor says publicly :

"At least one of you has red eyes."

the fact P becomes common knowledge.

Crucial point

The visitor does not change the world.
The visitor does not even give the islanders a new visible fact.
The visitor changes the *knowledge structure* :

from mutual knowledge to common knowledge.

A simple analogy

Suppose everyone in a classroom can see that it is raining outside.

Then probably everyone knows :

"It is raining."

But if the teacher says aloud :

"It is raining."

then something stronger happens :

- everyone knows it,
- everyone knows that everyone heard it,
- everyone knows that everyone knows that everyone heard it,
- and so on.

Moral

A public announcement can turn an obvious fact into *common knowledge*.

Take-away

Main lesson

The puzzle works not because the visitor reveals a hidden fact, but because the visitor makes a fact public at every level.

$$\text{mutual knowledge} \neq \text{common knowledge}$$

And in this puzzle, that difference changes everything.

A real-life example : traffic laws

Suppose you want to cross the street at a green light.

You feel safe only because you expect drivers to stop at red lights.

But why do you expect that ?

Not just because the rule exists, but because it is publicly known.

Let

$$P = \text{"Crossing a red light is forbidden and has serious consequences."}$$

What matters is not only that P is true, but that everyone knows it.

Why common knowledge matters

To cross safely, you want more than :

"I know the rule."

You also want :

- drivers know the rule ;
- pedestrians know the rule ;
- drivers know that pedestrians know it ;
- pedestrians know that drivers know it.

Key idea

Traffic works because the rule is not just known — it is common knowledge.

That is why a green light can serve as a reliable signal.

Game theory example : attack at dawn ?

Two generals must decide whether to attack.

- If **both** attack, they win.
- If only **one** attacks, that army is destroyed.
- So attacking is safe only if each general is sure the other will attack too.

They communicate by messengers, but a messenger may be lost.

The problem

General A sends : *"Attack at dawn."*
Even if General B receives the message, A does not know that B received it.

So A waits for a confirmation.

Why confirmations never seem enough

Suppose B sends back :

"I received your message."

Now A knows that B knows.

But a new problem appears :

- Does B know that A received the confirmation ?
- If A sends another confirmation, does A know that B received *that* one ?

Key idea

With unreliable communication, one confirmation is not enough, two are not enough, and in fact no finite number is enough to create common knowledge.

This also happens in modern communication systems :

email → reply → acknowledgement → . . .

A dangerous strategy

A strategic example : nuclear deterrence

During the Cold War, two rival states each had nuclear weapons.

- If one side launches a nuclear attack,
- the other side can retaliate,
- so both sides would suffer catastrophic destruction.

Mutually Assured Destruction (MAD)

The idea is that a nuclear attack is deterred because retaliation is expected.

But for deterrence to work, it is not enough that retaliation is merely *possible*.
It must also be publicly known to be certain.

Why common knowledge matters for deterrence

Suppose country A is considering an attack on country B.

Deterrence is stable only if :

- A knows that B can retaliate ;
- A knows that B is willing to retaliate ;
- B knows that A knows this ;
- and both sides know that both sides know it.

Key idea

Deterrence depends on retaliation being common knowledge, not just secret military information.

That is why states make military capabilities visible :

- official doctrine,
- public declarations,
- missile tests,
- visible arsenals.

What if the one is **not certain** of some knowledge, but only guesses it with some estimated *probability* ?

Deep subject into **Game theory** :

- (coordinated or uncoordinated) competitions
- elections
- pricing
- politics
- war and peace
- ...

To confess or not to confess, that is the question.

A probabilistic model : when do you confess ?

Suppose Alice is deciding whether to confess her feelings to Bob.

$$p = \text{Probability(Bob feels the same way)}.$$

Assume Alice assigns the following payoffs :

- $+10$ if she confesses and Bob feels the same : Great ;
- -6 if she confesses and Bob does no : Awkward ;
- 0 if she stays silent : nothing happens (but can regret it for ever).

So the expected payoff of confessing is

$$10p - 6(1-p) = 16p - 6.$$

Alice should confess when $16p - 6 > 0 \quad \Longleftrightarrow \quad p > \frac{3}{8} = 0.375.$

Idea

Confessing becomes rational only when the estimated probability is high enough.

Why dating matters

Dating can be seen as a way to collect evidence.

For example :

- spending time together,
- warm messages,
- attention and care,
- making future plans.

These signals can increase Alice's estimate :

$$p_0 \to p_1 \to p_2 \to p_3 \to \cdots$$

For instance,

$$0.10 \to 0.25 \to 0.45 \to 0.70$$

and once p passes the threshold 0.375, confessing becomes rational.

Interpretation

Dating/communication is a process of updating probabilities.

From probability update to public announcement

Before anyone speaks, both people may already have strong feelings.

But what matters is not only :

"I think you like me."

It also matters whether this becomes shared knowledge.

When Alice estimates the probability bigger than the threshold, she says :

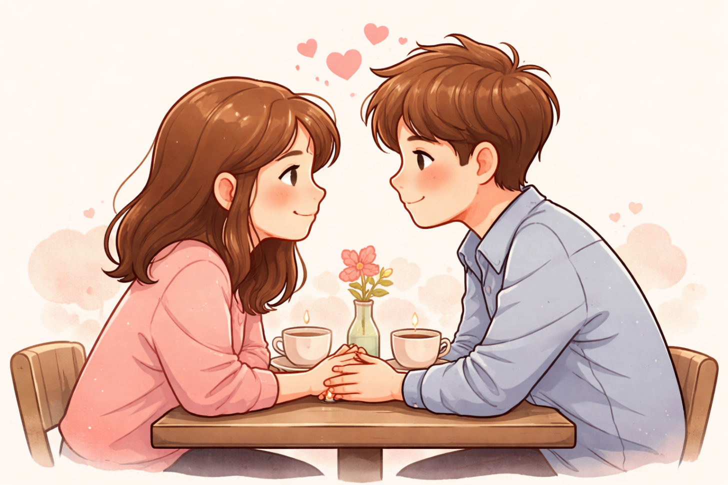
"I really like you."

the situation changes suddenly.

- Bob receives new information ;
- Bob updates his estimate ;
- the feelings are no longer just private — they become public.

Take-away

Dating is probability update ; confession is a public announcement.



Thank you !