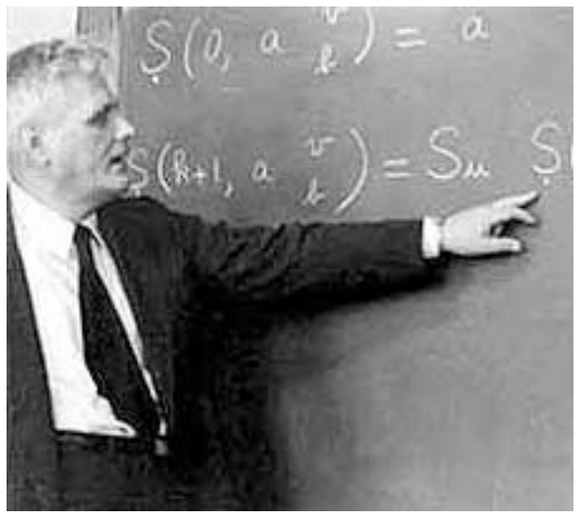
The paradox of knowability, the knower, and the believer

Last time, when discussing the surprise exam paradox, we discussed the possibility that some claims could be true, but not knowable by certain individuals - an example is that The Announcement might well be true, but can't be known to be true by students in the class (so long as they are sufficiently good at logic, and know the relevant facts about what they know).

A different paradox - the paradox of knowability - aims to show something stronger: that there are some truths which are **in principle unknowable**; not knowable by anyone at all.

The paradox is due to the great 20th century logician Alonzo Church, who communicated it to Frederic Fitch - who was the first to publish a version of the paradox, in 1963.



The paradox can be expressed as a reductio of the claim that every truth can be known (as before, we use "K" to abbreviate "it is known that"):

> **Knowability**: Every proposition P is such that if P is true, then it is possible that KP.

Intuitively, this claim seems quite plausible. Of course, perhaps there are some truths which we, given our cognitive limitations, can't know; but it seems quite plausible that we can imagine increasingly intelligent versions of ourselves who would, ultimately, be able to understand and know these truths.

But now consider the quite plausible claim that there are some truths which, even if knowable, are not actually known

> **Some Unknown**: There is at least one proposition P* such that (P* & ¬ KP*) - i.e., there is at least one proposition which is true but not known.

But Knowability is supposed to apply to every proposition; so it must also apply to the following proposition:

P* & ¬ KP*.

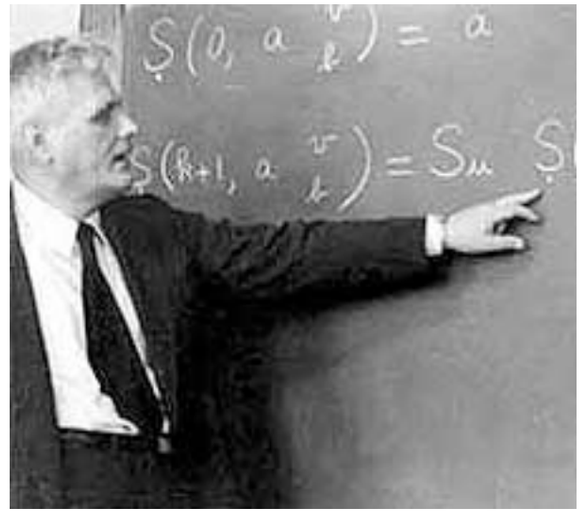
Hence the following must be true:

It is possible that K(P* & ¬ KP*)

But this result is problematic, because along with two very plausible principles, it implies a contradiction.

Knowability: Every proposition P is such that if P is true, then it is possible that KP.

Some Unknown: There is at least one proposition P* such that (P* & ¬ KP*) - i.e., there is at least one proposition which is true but not known.



But Knowability is supposed to apply to every proposition; so it must also apply to the following proposition:

P* & ¬ KP*.

Hence the following must be true:

It is possible that K(P* & ¬ KP*)

But this result is problematic, because along with two very plausible principles, it implies a contradiction.

The first principle is that knowledge of a conjunction - an “and” sentence - implies knowledge of both of its conjuncts. To see the plausibility of this, just think of some examples. As you will quickly find, whenever someone knows that P & Q, they also know that P, and know that Q. But this, plus the above result, implies:

It is possible that KP* & K(¬ KP*)

The next principle is one we discussed last time: the principle that knowledge implies truth. In our framework, this means that KP always implies P. Applying this rule, we get:

It is possible that KP* & ¬ KP*.

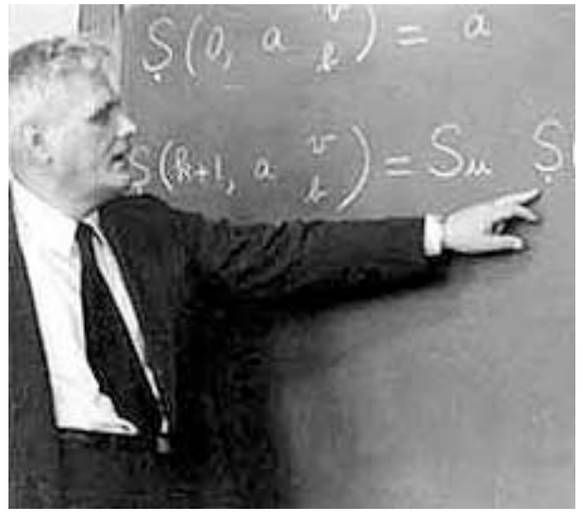
This says that it is possible that our proposition P* is both known and not known - but this is **not** possible, because it is a contradiction, and it is impossible for a contradiction to be true.

Hence, assuming the truth of our two principles, we seem to have a reductio of the principle of Knowability.

Does this succeed in showing that there are some truths which are, in principle, unknowable? Is this mysterious?

Knowability: Every proposition P is such that if P is true, then it is possible that KP.

Some Unknown: There is at least one proposition P* such that (P* & ¬ KP*) - i.e., there is at least one proposition which is true but not known.



P* & ¬ KP*.

It is possible that K(P* & ¬ KP*)

It is possible that KP* & K(¬ KP*)

It is possible that KP* & ¬ KP*.

This says that it is possible that our proposition P* is both known and not known - but this is **not** possible, because it is a contradiction, and it is impossible for a contradiction to be true.

Hence, assuming the truth of our two principles, we seem to have a reductio of the principle of Knowability.

Does this succeed in showing that there are some truths which are, in principle, unknowable? Is this mysterious?

Should this result worry the theist, since the existence of in-principle unknowable truths would presumably rule out the existence of an omniscient being?

Many have thought that this paradox rules out certain forms of philosophical idealism, or anti-realism, or relativism, which hold that reality is not completely independent of minds and their mental activity, or that truths are only true relative to minds and their mental activity. It seems quite plausible that if some such view were true, then there could be no truths which were in principle unknowable - for, if there were, wouldn't they have to be part of a reality which was in no way dependent on the mind for its existence? (Or perhaps it shows that the idealist, or relativist, should also be a theist, who can then escape the argument by denying that there are any unknown truths.)

Let's now turn to a quite different example of an unknowable statement: that involved in the paradox of **the Knower**.

Let's now turn to a quite different example of an unknowable statement: that involved in the paradox of **the Knower**.

Imagine that, when making The Announcement discussed in the surprise exam paradox, I concluded with these words:

> **The Conclusion (TC)**
>
> .. and, what's more, you know that this very announcement (TC, for short) is not true.

Consider now statement TC itself. It seems that we can give a proof that if TC is true, it is also false:

1. TC is true.	assumed for conditional proof
2. It is known that TC is not true.	1, plus the definition of TC
3. TC is not true.	knowledge implies truth

This is, in form, like van Inwagen's consequence argument: it is a conditional proof in which we assume P as a premise, show that it implies Q, and conclude from this that the conditional statement if P, then Q is true. From the above conditional proof we can thus conclude:

If TC is true, then TC is not true.

From which it follows that

TC is not true.

since, after all, TC must be either (1) true or (2) not true. If (1) then, given the above result, it is not true. And if (2) it is not true. Hence it is not true.

Now, as mentioned above, knowledge implies truth; hence, if TC is not true, it follows that we do not know TC. That is:

$\neg K(TC)$

The Conclusion (TC)

.. and, what's more, you know that this very announcement (TC, for short) is not true.

Consider now statement TC itself. It seems that we can give a proof that if TC is true, it is also false:

1. TC is true.	assumed for conditional proof
2. It is known that TC is not true.	1, plus the definition of TC
3. TC is not true.	knowledge implies truth

From which it follows that

TC is not true.

since, after all, TC must be either (1) true or (2) not true. If (1) then, given the above result, it is not true. And if (2) it is not true. Hence it is not true.

Now, as mentioned above, knowledge implies truth; hence, if TC is not true, it follows that we do not know TC. That is:

¬ K(TC)

So far, no big surprise.

But note something else about our situation: we have proven that TC is not true. But it seems clear that **if we have proven something, then we know it to be true** - indeed, proof seems to be the **surest** way to knowledge. But then, if we can prove that TC is not true, it follows that:

K(¬ TC)

But, of course,

TC = K(¬ TC).

Hence if we have shown that K(¬ TC), we have also shown

TC.

But the result of our conditional proof above was that TC is **not** true. What could be going on?

We have arguments for the contrary claims that TC is not true, and that it is true; let's try to get clear on how these two arguments work.

Conditional proof that if TC is true, then TC is not true

1. TC is true.	assumed for conditional proof
2. It is known that TC is not true.	1, plus the definition of TC
3. TC is not true.	knowledge -> truth
C1. If TC is true, then TC is not true	1-3, conditional proof

The Conclusion (TC)

.. and, what's more, you know that this very announcement (TC, for short) is not true.

Proof that TC is not true

6. TC is true or not true.	premise
7. If TC is not true, then TC is not true.	premise
8. If TC is true, then TC is not true.	C1
C2. TC is not true.	6,7,8

Proof that TC is true

9. We have proven that TC is not true.	6-C2
10. K(TC is not true)	9, proof -> knowledge
11. TC = K(TC is not true).	definition of TC
C3. TC	10,11

Both of these proofs look pretty hard to reject. But **something** must be wrong with them, since C2 contradicts C3..

It is tempting to say that the problem here can be solved by saying that the announcement is just **nonsense**. But there are two worries about the idea that we can respond to the paradox in that way.

First, it is not obvious that it is nonsense - and even if it is, we can re-create the paradox using statements which do seem to make sense. Imagine that you witness the following two statements by two people, A and B (you can imagine that they are in separate rooms, and their aim is to make a prediction about the statement made by the other):

> **A**: The next thing that B says is going to be something that you know is not true.

Then, a few moments later:

> **B**: The last thing that A said is true.

These certainly seem to make sense; so we can ask whether you know what B said or not. And, as above, either way of answering this question leads to a contradiction.

For suppose you do know that what B said is true. Then it follows that what A said is true, which in turn implies that you don't know, after all, that what B said is true.

Suppose instead that you do not know that what B said is true. This implies that what A said is true, which in turn implies that what B said is true --- and the above line of reasoning would seem to put you in a position to know this.

Conditional proof that if TC is true, then TC is not true

1. TC is true.	assumed for conditional proof
2. It is known that TC is not true.	1, plus the definition of TC
3. TC is not true.	knowledge -> truth
C1. If TC is true, then TC is not true	1-3, conditional proof

The Conclusion (TC)

.. and, what's more, you know that this very announcement (TC, for short) is not true.

Proof that TC is not true

6. TC is true or not true.	premise
7. If TC is not true, then TC is not true.	premise
8. If TC is true, then TC is not true.	C1
C2. TC is not true.	6,7,8

Proof that TC is true

9. We have proven that TC is not true.	6-C2
10. K(TC is not true)	9, proof -> knowledge
11. TC = K(TC is not true).	definition of TC
C3. TC	10,11

Both of these proofs look pretty hard to reject. But **something** must be wrong with them, since C2 contradicts C3..

It is tempting to say that the problem here can be solved by saying that the announcement is just **nonsense**. But there are two worries about the idea that we can respond to the paradox in that way.

A second, and more fundamental, problem is that it's not clear how the claim that TC is nonsense could, even if true, help with this paradox. What premise in the above arguments would it show to be false?

One might think: premise 6. But if TC is nonsense, then it seems to follow that TC is not true - nonsense statements, after all, are never true. But this would make 6 true.

Conditional proof that if TC is true, then TC is not true

1. TC is true.	assumed for conditional proof
2. It is known that TC is not true.	1, plus the definition of TC
3. TC is not true.	knowledge -> truth
C1. If TC is true, then TC is not true	1-3, conditional proof

The Conclusion (TC)

.. and, what's more, you know that this very announcement (TC, for short) is not true.

Proof that TC is not true

6. TC is true or not true.	premise
7. If TC is not true, then TC is not true.	premise
8. If TC is true, then TC is not true.	C1
C2. TC is not true.	6,7,8

Proof that TC is true

9. We have proven that TC is not true.	6-C2
10. K(TC is not true)	9, proof -> knowledge
11. TC = K(TC is not true).	definition of TC
C3. TC	10,11

Both of these proofs look pretty hard to reject. But **something** must be wrong with them, since together they imply a contradiction.

A plausible thought is that, as Sainsbury says, the paradox has essentially to do with the fact that knowledge implies truth. This assumption is, after all, used in both of our proofs. We can't, however, reasonably reject this assumption; but perhaps we **can** say that the paradox here arises not from the nature of knowledge, but rather from the nature of truth.

After all, it seems that we can generate a very similar looking paradox using just the notion of truth, without bringing in knowledge at all:

L1. L1 is not true.

This is the **Liar paradox**. It, like TC, seems to lead to a contradiction. For consider: L1 is either true or not true. If it is not true, then L1 is just the say it says it is - so it must be true. But if L1 is true, then L1 must be the way it says it is: namely, untrue. So if it is not true, then it is true, and if it is true, it is not true. So it must be neither true nor not true - but this certainly sounds like a contradiction, since it sounds like we are saying that it is both not true and not not true - which is definitely a contradiction.

The Liar is, arguably, the hardest and most fundamental paradox that we will be discussing this semester. But because of the similarities between the Knower and the Liar - such as the fact that both "say of themselves" something that implies that they are not true - perhaps the right thing to say about the Knower is just that in order to solve it, we need to solve the Liar; and whatever ends up being the right solution to the Liar will also give us a solution to the Knower.

This is a very natural thought. But, if it is correct, then we should expect that we will **not** be able to raise a similar paradox using mental states such as belief which, unlike knowledge, do not imply truth. Unfortunately, as it turns out, we can.

This is a very natural thought. But, if it is correct, then we should expect that we will **not** be able to raise a similar paradox using mental states such as belief which, unlike knowledge, do not imply truth. Unfortunately, as it turns out, we can.

Suppose that I finish The Announcement not with (TC), but with the following modified conclusion:

The Conclusion, Part 2 (TC2)

.. and, what's more, none of you believe this very announcement (TC2, for short).

Now let's suppose a few things about you. First, let's suppose that your logical abilities are adequate for you to carry out the logical inferences described below; and, second, let's suppose that you form beliefs according to the following rule: if you can see that something is true, you will believe it; and if you can see that it is untrue, you won't believe it. (Arguably, it is impossible for someone capable of forming beliefs not to follow this rule; it does not seem that we have a choice about whether to believe things that we take to be true, and one cannot believe something at will which one can see to be false.)

Let's use "B" to abbreviate "believes" much as we used "K" to abbreviate "knows." Then let's ask: do you believe TC2?

Let's suppose that you do not. Then you can see that TC2 is true. But then, given the second rule of rationality above, it follows that B(TC2).

So suppose that B(TC2). Then TC2 is false. But because you see this, given the second rule above - that you never believe something which you can see to be false - you don't believe TC2. So, ¬ B(TC2).

So, if B(TC2) then ¬ B(TC2), and if ¬ B(TC2), then B(TC2) - which is a contradiction, just as in the case of the original announcement.

We cannot explain this in terms of belief implying truth because, of course, belief does **not** imply truth. But perhaps the contradiction produced by the paradox of the Believer can be given a related explanation. Deriving the contradiction, after all, made essential use of rules which link belief to truth - namely, the rule that known truths will be believed, and known falsehoods not believed. It does not seem possible to recreate the paradox using mental states which are not governed by any rules to do with truth. So perhaps the Believer, like the Knower, can ultimately only be solved by solving the Liar.